PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N

A2

(11) International Publication Number: WO 95/33818

(43) International Publication Date: 14 December 1995 (14.12.95)

(21) International Application Number: PCT/IB95/00414

(22) International Filing Date: 30 May 1995 (30.05.95)

(30) Priority Data:
08/258,261 8 June 1994 (08.06.94) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/258,261 (CIP)
Filed on 8 June 1994 (08.06.94)

(71) Applicant (*for all designated States except US*): CIBA-GEIGY AG [CH/CH]; Klybeckstrasse 141, CH-4002 Basle (CH).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): LIGON, James, Madison [US/CH]; Oetlingerstrasse 2, CH-4057 Basle (CH). SCHUPP, Thomas [CH/CH]; Fröschmattweg 5, CH-4313 Möhlin (CH). BECK, James, Joseph [US/US]; 114 Ripley Court, Cary, NC 27513 (US). HILL, Dwight, Steven [US/US]; 311 Melanie Lane, Cary, NC 27511 (US). RYALS, John, Andrew [US/US]; 14 Sanderling Court, Durham, NC 27713 (US). GAFFNEY, Thomas, Deane [US/US]; 125 Tradescent Road, Chapel Hill, NC 27514 (US). LAM, Stephen, Ting [US/US]; 8900 Jeanew Court, Raleigh, NC 27613 (US). HAMMER, Philip, Eugene [US/US]; 212 A Waldo Street, Cary, NC 27511 (US). UKNES, Scott, Joseph [US/US]; 1003 Pinedale Drive, Apex, NC 27502 (US).

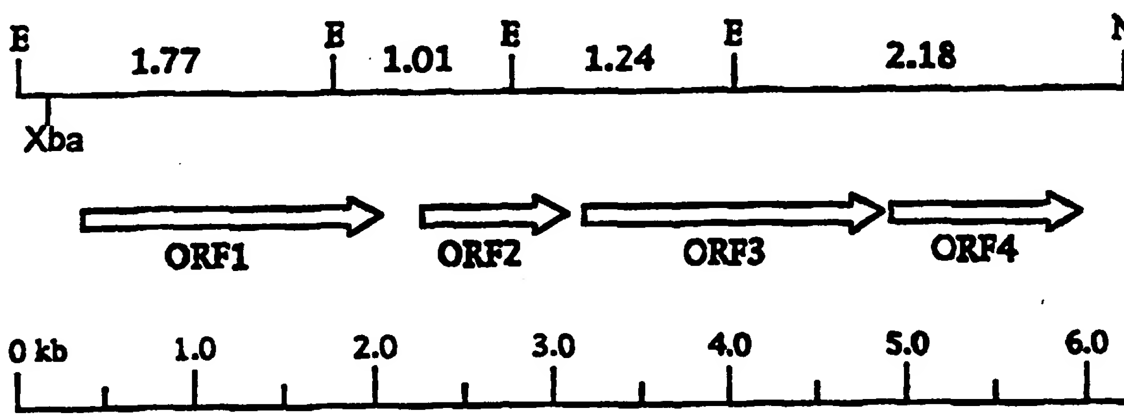
(74) Common Representative: CIBA-GEIGY AG; Patentabteilung, Klybeckstrasse 141, CH-4002 Basle (CH).

(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GENES FOR THE SYNTHESIS OF ANTIPATHOGENIC SUBSTANCES

Prn Gene Region of MOCG134



(57) Abstract

The present invention is directed to the production of an antipathogenic substance (APS) in a host via recombinant expression of the polypeptides needed to biologically synthesize the APS. Genes encoding polypeptides necessary to produce particular antipathogenic substances are provided, along with methods for identifying and isolating genes needed to recombinantly biosynthesize any desired APS. The cloned genes may be transformed and expressed in a desired host organisms to produce the APS according to the invention for a variety of purposes, including protecting the host from a pathogen, developing the host as a biocontrol agent, and producing large uniform amounts of the APS.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

GENES FOR THE SYNTHESIS OF ANTIPATHOGENIC SUBSTANCES

The present invention relates generally to the protection of host organisms against pathogens, and more particularly to the protection of plants against phytopathogens. In one aspect it provides transgenic plants which have enhanced resistance to phytopathogens and biocontrol organisms with enhanced biocontrol properties. It further provides methods for protecting plants against phytopathogens and methods for the production of antipathogenic substances.

Plants routinely become infected by fungi and bacteria, and many microbial species have evolved to utilize the different niches provided by the growing plant. Some phytopathogens have evolved to infect foliar surfaces and are spread through the air, from plant-to-plant contact or by various vectors, whereas other phytopathogens are soil-borne and preferentially infect roots and newly germinated seedlings. In addition to infection by fungi and bacteria, many plant diseases are caused by nematodes which are soil-borne and infect roots, typically causing serious damage when the same crop species is cultivated for successive years on the same area of ground.

Plant diseases cause considerable crop loss from year to year resulting both in economic hardship to farmers and nutritional deprivation for local populations in many parts of the world. The widespread use of fungicides has provided considerable security against phytopathogen attack, but despite $1 billion worth of expenditure on fungicides, worldwide crop losses amounted to approximately 10% of crop value in 1981 (James, Seed Sci. & Technol. 9: 679-685 (1981). The severity of the destructive process of disease depends on the aggressiveness of the phytopathogen and the response of the host, and one aim of most plant breeding programs is to increase the resistance of host plants to disease. Novel gene sources and combinations developed for resistance to disease have typically only had a limited period of successful use in many crop-pathogen systems due to the rapid evolution of phytopathogens to overcome resistance genes. In addition, there are several documented cases of the evolution of fungal strains which are resistant to particular fungicides. As early as 1981, Fletcher and Wolfe (Proc. 1981 Brit. Crop Prot. Conf. (1981))

contended that 24% of the powdery mildew populations from spring barley, and 53% from winter barley showed considerable variation in response to the fungicide triadimenol and that the distribution of these populations varied between barley varieties with the most susceptible variety also giving the highest incidence of less susceptible fungal types. Similar variation in the sensitivity of fungi to fungicides has been documented for wheat mildew (also to triadimenol), *Botrytis* (to benomyl), *Pyrenophora* (to organomercury), *Pseudocercosporella* (to MBC-type fungicides) and *Mycosphaerella fijiensis* to triazoles to mention just a few (Jones and Clifford; Cereal Diseases, John Wiley, 1983). Diseases caused by nematodes have also been controlled successfully by pesticide application. Whereas most fungicides are relatively harmless to mammals and the problems with their use lie in the development of resistance in target fungi, the major problem associated with the use of nematicides is their relatively high toxicity to mammals. Most nematicides used to control soil nematodes are of the carbamate, organochlorine or organophosphorous groups and must be applied to the soil with particular care.

In some crop species, the use of biocontrol organisms has been developed as a further alternative to protect crops. Biocontrol organisms have the advantage of being able to colonize and protect parts of the plant inaccessible to conventional fungicides. This practice developed from the recognition that crops grown in some soils are naturally resistant to certain fungal phytopathogens and that the suppressive nature of these soils is lost by autoclaving. Furthermore, it was recognized that soils which are conducive to the development of certain diseases could be rendered suppressive by the addition of small quantities of soil from a suppressive field (Scher *et al.* Phytopathology 70: 412-417 (1980). Subsequent research demonstrated that root colonizing bacteria were responsible for this phenomenon, now known as biological disease control (Baker *et al.* Biological Control of Plant Pathogens, Freeman Press, San Francisco, 1974). In many cases, the most efficient strains of biological disease controlling bacteria are of the species *Pseudomonas fluorescens* (Weller *et al.* Phytopathology 73: 463-469 (1983); Kloepper *et al.* Phytopathology 71: 1020-1024 (1981)). Important plant pathogens that have been effectively controlled by seed inoculation with these bacteria include *Gaemannomyces graminis*, the causative agent of take-all in wheat (Cook *et al.* Soil Biol. Biochem 8: 269-273 (1976)) and the *Pythium* and *Rhizoctonia* phytopathogens involved in damping off of cotton (Howell *et al.* Phytopathology 69: 480-482 (1979)). Several biological disease controlling

Pseudomonas strains produce antibiotics which inhibit the growth of fungal phytopathogens (Howell *et al.* Phytopathology 69: 480-482 (1979); Howell *et al.* Phytopathology 70: 712-715 (1980)) and these have been implicated in the control of fungal phytopathogens in the rhizosphere. Although biocontrol was initially believed to have considerable promise as a method of widespread application for disease control, it has found application mainly in the environment of glasshouse crops where its utility in controlling soil-borne phytopathogens is best suited for success. Large scale field application of naturally occurring microorganisms has not proven possible due to constraints of microorganism production (they are often slow growing), distribution (they are often short lived) and cost (the result of both these problems). In addition, the success of biocontrol approaches is also largely limited by the identification of naturally occurring strains which may have a limited spectrum of efficacy. Some initial approaches have also been taken to control nematode phytopathogens using biocontrol organisms. Although these approaches are still exploratory, some *Streptomyces* species have been reported to control the root knot nematode (*Meliodogyne* spp.) (WO 93/18135 to Research Corporation Technology), and toxins from some *Bacillus thuringiensis* strains (such as *israeliensis*) have been shown to have broad anti-nematode activity and spore or bacillus preparations may thus provide suitable biocontrol opportunities (EP 0 352 052 to Mycogen, WO 93/19604 to Research Corporation Technologies).

The traditional methods of protecting crops against disease, including plant breeding for disease resistance, the continued development of fungicides, and more recently, the identification of biocontrol organisms, have all met with success. It is apparent, however, that scientists must constantly be in search of new methods with which to protect crops against disease. This invention provides novel methods for the protection of plants against phytopathogens.

The present invention reveals the genetic basis for substances produced by particular microorganisms via a multi-gene biosynthetic pathway which have a deleterious effect on the multiplication or growth of plant pathogens. These substances include carbohydrate containing antibiotics such as aminoglycosides, peptide antibiotics, nucleoside derivatives and other heterocyclic antibiotics containing nitrogen and/or oxygen, polyketides, macrocyclic lactones, and quinones.

The invention provides the entire set of genes required for recombinant production of particular antipathogenic substances in a host organism. It further provides methods for the manipulation of APS gene sequences for their expression in transgenic plants. The transgenic plants thus modified have enhanced resistance to attack by phytopathogens. The invention provides methods for the cellular targeting of APS gene products so as to ensure that the gene products have appropriate spatial localization for the availability of the required substrate/s. Further provided are methods for the enhancement of throughput through the APS metabolic pathway by overexpression and overproduction of genes encoding substrate precursors.

The invention further provides a novel method for the identification and isolation of the genes involved in the biosynthesis of any particular APS in a host organism.
The invention also describes improved biocontrol strains which produce heterologous APSs and which are efficacious in controlling soil-borne and seedling phytopathogens outside the usual range of the host.

Thus, the invention provides methods for disease control. These methods involve the use of transgenic plants expressing APS biosynthetic genes and the use of biocontrol agents expressing APS genes.

The invention further provides methods for the production of APSs in quantities large enough to enable their isolation and use in agricultural formulations. A specific advantage of these production methods is the uniform chirality of the molecules produced; production in transgenic organisms avoids the generation of populations of racemic mixtures, within which some enantiomers may have reduced activity.

DEFINITIONS

As used in the present application, the following terms have the meanings set out below.
Antipathogenic Substance: A substance which requires one or more nonendogenous enzymatic activities foreign to a plant to be produced in a host where it does not naturally occur, which substance has a deleterious effect on the multiplication or growth of a pathogen (i.e. pathogen). By " nonendogenous enzymatic activities" is meant enzymatic

activities that do not naturally occur in the host where the antipathogenic substance does not naturally occur. A pathogen may be a fungus, bacteria, nematode, virus, viroid, insect or combination thereof, and may be the direct or indirect causal agent of disease in the host organism. An antipathogenic substance can prevent the multiplication or growth of a phytopathogen or can kill a phytopathogen. An antipathogenic substance may be synthesized from a substrate which naturally occurs in the host. Alternatively, an antipathogenic substance may be synthesized from a substrate that is provided to the host along with the necessary nonendogenous enzymatic activities. An antipathogenic substance may be a carbohydrate containing antibiotic, a peptide antibiotic, a heterocyclic antibiotic containing nitrogen, a heterocyclic antibiotic containing oxygen, a heterocyclic antibiotic containing nitrogen and oxygen, a polyketide, a macrocyclic lactone, and a quinone. Antipathogenic substance is abbreviated as "APS" throughout the text of this application.

Anti-phytopathogenic substance: An antipathogenic substance as herein defined which has a deleterious effect on the multiplication or growth of a plant pathogen (i.e.phytopathogen).

Biocontrol agent: An organism which is capable of affecting the growth of a pathogen such that the ability of the pathogen to cause a disease is reduced. Biocontrol agents for plants include microorganisms which are capable of colonizing plants or the rhizosphere. Such biocontrol agents include gram-negative microorganisms such as *Pseudomonas, Enterobacter* and *Serratia*, the gram-positive microorganism *Bacillus* and the fungi *Trichoderma* and *Gliocladium*. Organisms may act as biocontrol agents in their native state or when they are genetically engineered according to the invention.

Pathogen: Any organism which causes a deleterious effect on a selected host under appropriate conditions. Within the scope of this invention the term pathogen is intended to include fungi, bacteria, nematodes, viruses, viroids and insects.

Promoter or Regulatory DNA Sequence: An untranslated DNA sequence which assists in, enhances, or otherwise affects the transcription, translation or expression of an associated structural DNA sequence which codes for a protein or other DNA product. The promoter

DNA sequence is usually located at the 5' end of a translated DNA sequence, typically between 20 and 100 nucleotides from the 5' end of the translation start site.

Coding DNA Sequence: A DNA sequence that is translated in an organism to produce a protein.

Operably Linked to/Associated With: Two DNA sequences which are "associated" or "operably linked" are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that codes for an RNA or a protein if the two sequences are operably linked, or situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

Chimeric Construction/Fusion DNA Sequence: A recombinant DNA sequence in which a promoter or regulatory DNA sequence is operably linked to, or associated with, a DNA sequence that codes for an mRNA or which is expressed as a protein, such that the regulator DNA sequence is able to regulate transcription or expression of the associated DNA sequence. The regulator DNA sequence of the chimeric construction is not normally operably linked to the associated DNA sequence as found in nature. The terms "heterologous" or "non-cognate" are used to indicate a recombinant DNA sequence in which the promoter or regulator DNA sequence and the associated DNA sequence are isolated from organisms of different species or genera.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Restriction map of the cosmid clone pCIB169 from *Pseudomonas fluorescens* carrying the pyrrolnitrin biosynthetic gene region. Restricition sites of the enzymes EcoRI, HindIII, KpnI, NotI, SphI, and XbaI as well as nucleotide positions in kbp are indicated.

Figure 2: Functional Map of the Pyrrolnitrin Gene Region of MOCG134 indicating insertion points of 30 independent Tn5 insertions along the length of pCIB169 for the identification of the genes for pyrrolnitrin biosynthesis. EcoRI restriction sites are

designated with E, NotI sites with N. The effect of a Tn5 insertion on prn production is designated with either + or -, wherein + indicates a prn producer and - a prn non-producer.

Figure 3: Restriction map of the 9.7 kb MOCG134 Prn gene region of clone pCIB169 involved in pyrrolnitrin biosynthesis. EcoRI restriction sites are designated with E, NotI sites with N, and HindIII sites with H. Nucleotide positions are indicated in kbp.

Figure 4: Location of various subclones derived from pCIB169 isolated for sequence determination purposes.

Figure 5: Localization of the four open reading frames (ORFs 1-4) responsible for pyrrolnitrin biosynthesis in strain MOCG134 on the ~6 kb *XbaI/NotI* fragment of pCIB169 comprising the Prn gene region.

Figure 6: Location of the fragments deleted in ORFs 1-4 in the pyrrolnitrin gene cluster of MOCG134. Deleted fragments are indicated as filled boxes.

Figure 7: Restriction map of the cosmid clone p98/1 from *Sorangium cellulosum* carrying the soraphen biosynthetic gene region. The top line depicts the restriction map of p98/1 and shows the position of restriction sites and their distance from the left edge in kilobases. Restriction sites shown include: B, Bam HI; Bg Bg1 II; E, Eco RI; H, Hind III; Pv, Pvu I; Sm, Sma I. The boxes below the restriction map depict the location of the biosynthetic modules. The activity domains within each module are designated as follows: β-ketoacylsynthase (KS), Acyltransferase (AT), Ketoreductase (KR), Acyl Carrier Protein (ACP), Dehydratase (DH), Enoyl reductase (ER), and Thioesterase (TE).

Figure 8: Construction of pCIB132 from pSUP2021.

Figure 9: Restriction endonuclease map of the phenazine biosynthetic gene cluster contained on a 5.7 kb *EcoRI-HindIII* fragment. Orientation and approximate positions of the six open reading frames are presented below the restriction map. ORF1, which is not entirely present within the 5.7 kb fragment, encodes a product with significant homology to plant DAHP synthases. ORF2 (0.65 kb), ORF3 (0.75 kb), and ORF4 (1.15 kb) have domains homologous to isochorismatase, anthranilate synthase large subunit, and anthranilate synthase small subunit, respectively. ORF5 (0.7 kb) demonstrates no homology with database sequences. The ORF6 (0.65 kb) product has end to end homology with the gene encoding pyridoxine 5'-phosphate oxidase in E. coli.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

SEQ ID NO:1: ....Sequence of the Pyrrolnitrin Gene Cluster
SEQ ID NO:2: ....Protein sequence for ORF1 of pyrrolnitrin gene cluster
SEQ ID NO:3: ....Protein sequence for ORF2 of pyrrolnitrin gene cluster
SEQ ID NO:4: ....Protein sequence for ORF3 of pyrrolnitrin gene cluster
SEQ ID NO:5: ....Protein sequence for ORF4 of pyrrolnitrin gene cluster
SEQ ID NO:6: ....Sequence of the Soraphen Gene Cluster
SEQ ID NO:7: ....Sequence of a Plant Consensus Translation Initiator (Clontech)
SEQ ID NO:8: ....Sequence of a Plant Consensus Translation Initiator (Joshi)
SEQ ID NO:9: ....Sequence of an Oligonucleotide for Use in a Molecular Adaptor
SEQ ID NO:10: ....Sequence of an Oligonucleotide for Use in a Molecular Adaptor
SEQ ID NO:11: ....Sequence of an Oligonucleotide for Use in a Molecular Adaptor
SEQ ID NO:12: ....Sequence of an Oligonucleotide for Use in a Molecular Adaptor
SEQ ID NO:13: ....Sequence of an Oligonucleotide for Use in a Molecular Adaptor
SEQ ID NO:14: ....Sequence of an Oligonucleotide for Use in a Molecular Adaptor
SEQ ID NO:15: ....Oligonucleotide used to change restriction site
SEQ ID NO:16: ....Oligonucleotide used to change restriction site
SEQ ID NO:17: ....Sequence of the Phenazine Gene Cluster
SEQ ID NO:18: ....Protein sequence for phz1 from the phenazine gene cluster
SEQ ID NO:19: ....Protein sequence for phz2 from the phenazine gene cluster
SEQ ID NO:20: ....Protein sequence for phz3 from the phenazine gene cluster
SEQ ID NO:21: ....DNA sequence for phz4 of Phenazine gene cluster
SEQ ID NO:22: ....Protein sequence for phz4 from the phenazine gene cluster

DEPOSITS

Clone	Accession Number	Date of Deposit
pJL3	NRRL B-21254	May 20, 1994
p98/1	NRRL B-21255	May 20, 1994
pCIB169	NRRL B-21256	May 20, 1994
pCIB3350	NRRL B-21257	May 20, 1994
pCIB3351	NRRL B-21258	May 20, 1994

Production of Antipathogenic Substances by Microorganisms

Many organisms produce secondary metabolites and some of these inhibit the growth of other organisms. Since the discovery of penicillin, a large number of compounds with antibiotic activity have been identified, and the number continues to increase with ongoing screening efforts. Antibiotically active metabolites comprise a broad range of chemical structures. The most important include: aminoglycosides (*e.g.* streptomycin) and other carbohydrate containing antibiotics, peptide antibiotics (*e.g.* β-lactAPS, rhizocticin (*see* Rapp, C. *et al., Liebigs Ann. Chem.* : 655-661 (1988)), nucleoside derivatives (*e.g.* blasticidin S) and other heterocyclic antibiotics containing nitrogen (*e.g.* phenazine and pyrrolnitrin) and/or oxygen, polyketides (*e.g.* soraphen), macrocyclic lactones (*e.g.* erythromycin) and quinones (*e.g.* tetracycline).

Aminoglycosides and Other Carbohydrate Containing Antibiotics

The aminoglycosides are oligosaccharides consisting of an aminocyclohexanol moiety glycosidically linked to other amino sugars. Streptomycin, one of the best studied of the group, is produced by *Streptomyces griseus*. The biochemistry and biosynthesis of this compound is complex (for review see Mansouri *et al.* in: Genetics and Molecular Biology of Industrial Microorganisms (*ed.*: Hershberger *et al.*), American Society for Microbiology, Washington, D. C. pp 61-67 (1989)) and involves 25 to 30 genes, 19 of which have been analyzed so far (Retzlaff *et al.* in: Industrial Microorganisms: Basic and Applied Molecular Genetics (*ed.*: Baltz *et al.*), American Society for Microbiology, Washington, D. C. pp 183-194 (1993)). Streptomycin, and many other aminoglycosides, inhibits protein synthesis in the target organisms.

Peptide Antibiotics

Peptide antibiotics are classifiable into two groups: (1) those which are synthesized by enzyme systems without the participation of the ribosomal apparatus, and (2) those which require the ribosomally-mediated translation of an mRNA to provide the precursor of the antibiotic.

Non-Ribosomal Peptide Antibiotics are assembled by large, multifunctional enzymes which activate, modify, polymerize and in some cases cyclize the subunit amino acids, forming polypeptide chains. Other acids, such as aminoadipic acid, diaminobutyric acid,

diaminopropionic acid, dihydroxyamino acid, isoserine, dihydroxybenzoic acid, hydroxyisovaleric acid, (4R)-4-[(E)-2-butenyl]-4,N-dimethyl-L-threonine, and ornithine are also incorporated (Katz & Demain, Bacteriological Review 41: 449-474 (1977); Kleinkauf & von Dohren, Annual Review of Microbiology 41: 259-289 (1987)). The products are not encoded by any mRNA, and ribosomes do not directly participate in their synthesis. Peptide antibiotics synthesized non-ribosomally can in turn be grouped according to their general structures into linear, cyclic, lactone, branched cyclopeptide, and depsipeptide categories (Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990)). These different groups of antibiotics are produced by the action of modifying and cyclizing enzymes; the basic scheme of polymerization is common to them all. Non-ribosomally synthesized peptide antibiotics are produced by both bacteria and fungi, and include edeine, linear gramicidin, tyrocidine and gramicidin S from *Bacillus brevis*, mycobacillin from *Bacillus subtilis*, polymyxin from *Bacillus polymiyxa*, etamycin from *Streptomyces griseus*, echinomycin from *Streptomyces echinatus*, actinomycin from *Streptomyces clavuligerus*, enterochelin from *Escherichia coli*, gamma-(alpha-L-aminoadipyl)-L-cysteinyl-D-valine (ACV) from *Aspergillus nidulans*, alamethicine from *Trichoderma viride*, destruxin from *Metarhizium anisolpliae*, enniatin from *Fusarium oxysporum*, and beauvericin from *Beauveria bassiana*. Extensive functional and structural similarity exists between the prokaryotic and eukaryotic systems, suggesting a common origin for both. The activities of peptide antibiotics are similarly broad, toxic effects of different peptide antibiotics in animals, plants, bacteria, and fungi are known (Hansen, Annual Review of Microbiology 47: 535-564 (1993); Katz & Demain, Bacteriological Reviews 41: 449-474 (1977); Kleinkauf & von Dohren, Annual Review of Microbiology 41: 259-289 (1987); Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990); Kolter & Moreno, Annual Review of Microbiology 46: 141-163 (1992)).

Ribosomally-Synthesized Peptide Antibiotics are characterized by the existence of a structural gene for the antibiotic itself, which encodes a precursor that is modified by specific enzymes to create the mature molecule. The use of the general protein synthesis apparatus for peptide antibiotic synthesis opens up the possibility for much longer polymers to be made, although these peptide antibiotics are not necessarily very large. In addition to a structural gene, further genes are required for extracellular secretion and immunity, and these genes are believed to be located close to the structural gene, in most cases probably

on the same operon. Two major groups of peptide antibiotics made on ribosomes exist: those which contain the unusual amino acid lanthionine, and those which do not. Lanthionine-containing antibiotics (lantibiotics) are produced by gram-positive bacteria, including species of *Lactococcus*, *Staphylococcus*, *Streptococcus*, *Bacillus*, and *Streptomyces*. Linear lantibiotics (for example, nisin, subtilin, epidermin, and gallidermin), and circular lantibiotics (for example, duramycin and cinnamycin), are known (Hansen, Annual Review of Microbiology 47: 535-564 (1993); Kolter & Moreno, Annual Review of Microbiology 46: 141-163 (1992)). Lantibiotics often contain other characteristic modified residues such as dehydroalanine (DHA) and dehydrobutyrine (DHB), which are derived from the dehydration of serine and threonine, respectively. The reaction of a thiol from cysteine with DHA yields lanthionine, and with DHB yields β-methyllanthionine. Peptide antibiotics which do not contain lanthionine may contain other modifications, or they may consist only of the ordinary amino acids used in protein synthesis. Non-lanthionine-containing peptide antibiotics are produced by both gram-positive and gram-negative bacteria, including *Lactobacillus*, *Lactococcus*, *Pediococcus*, *Enterococcus*, and *Escherichia*. Antibiotics in this category include lactacins, lactocins, sakacin A, pediocins, diplococcin, lactococcins, and microcins (Hansen, *supra*; Kolter & Moreno, *supra*).

Nucleoside Derivatives and Other Heterocyclic Antibiotics Containing Nitrogen and/or Oxygen

These compounds all contain heterocyclic rings but are otherwise structurally diverse and, as illustrated in the following examples, have very different biological activities.

Polyoxins and Nikkomycins are nucleoside derivatives and structurally resemble UDP-N-acetylglucosamine, the substrate of chitin synthase. They have been identified as competitive inhibitors of chitin synthase (Gooday, in: Biochemistry of Cell Walls and Membranes in Fungi (ed.: Kuhn *et al.*), Springer-Verlag, Berlin p. 61 (1990)). The polyoxins are produced by *Streptomyces cacaoi* and the Nikkomycins are produced by *S. tendae*.

Phenazines are nitrogen-containing heterocyclic compounds with a common planar aromatic tricyclic structure. Over 50 naturally occurring phenazines have been identified, each differing in the substituent groups on the basic ring structure. This group of compounds are found produced in nature exclusively by bacteria, in particular

Streptomyces, *Sorangium*, and *Pseudomonas* (for review see Turner & Messenger, Advances in Microbiol Physiology 27: 211-275 (1986)). Recently, the phenazine biosynthetic genes of a *P. aureofaciens* strain has been isolated (Pierson & Thomashow MPMI 5: 330-339 (1992)). Because of their planar aromatic structure, it has been proposed that phenazines may form intercalative complexes with DNA (Hollstein & van Gemert, Biochemistry 10: 497 (1971)), and thereby interfere with DNA metabolism. The phenazine myxin was shown to intercalate DNA (Hollstein & Butler, Biochemistry 11: 1345 (1972)) and the phenazine lomofungin was shown to inhibit RNA synthesis in yeast (Cannon & Jiminez, Biochemical Journal 142: 457 (1974); Ruet *et al.*, Biochemistry 14: 4651 (1975)).

Pyrrolnitrin is a phenylpyrrole derivative with strong antibiotic activity and has been shown to inhibit a broad range of fungi (Homma *et al.*, Soil Biol. Biochem. 21: 723-728 (1989); Nishida *et al.*, J. Antibiot., ser A, 18: 211-219 (1965)). It was originally isolated from *Pseudomonas pyrrocinia* (Arima et al, J. Antibiot., ser. A, 18: 201-204 (1965)), and has since been isolated from several other *Pseudomonas* species and *Myxococcus* species (Gerth *et al.* J. Antibiot. 35: 1101-1103 (1982)). The compound has been reported to inhibit fungal respiratory electron transport (Tripathi & Gottlieb, J. Bacteriol. 100: 310-318 (1969)) and uncouple oxidative phosphorylation (Lambowitz & Slayman, J. Bacteriol. 112: 1020-1022 (1972)). It has also been proposed that pyrrolnitrin causes generalized lipoprotein membrane damage (Nose & Arima, J. Antibiot., ser A, 22: 135-143 (1969); Carlone & Scannerini, Mycopahtologia et Mycologia Applicata 53: 111-123 (1974)). Pyrrolnitrin is biosynthesized from tryptophan (Chang *et al.* J. Antibiot. 34: 555-566) and the biosynthetic genes from *P. fluorescens* have now been cloned (see Section C of examples). Thus, one embodiment of the present invention relates to an isolated DNA molecule encoding one or more polypeptides for the biosynthesis of pyrrolnitrin in a heterologous host, which molecule can be used to genetically engineer a host organism to express said antipathogenic substance. Other embodiments of the invention are the isolated polypeptides required for the biosynthesis of pyrrolnitrin.

Polyketide Synthases

Many antibiotics, in spite of the apparent structural diversity, share a common pattern of biosynthesis. The molecules are built up from two carbon building blocks, the β-carbon of which always carries a keto group, thus the name polyketide. The tremendous structural

diversity derives from the different lengths of the polyketide chain and the different side-chains introduced, either as part of the two carbon building blocks, or after the polyketide backbone is formed. The keto groups may also be reduced to hydroxyls or removed altogether. Each round of two carbon addition is carried out by a complex of enzymes called the polyketide synthases (PKS) in a manner similar to fatty acid biosynthesis. The biosynthetic genes for an increasing number of polyketide antibiotics have been isolated and sequenced. It is quite apparent that the PKS genes are structurally conserved. The encoded proteins generally fall into two types: type I proteins are polyfunctional, with several catalytic domains carrying out different enzymatic steps covalently linked together (e.g. PKS for erythromycin, soraphen, and avermectin (Joaua *et al.* Plasmid 28: 157-165 (1992); MacNeil *et al.* in: Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz *et al.*), American Society for Microbiology, Washington D. C. pp. 245-256 (1993)); whereas type II proteins are monofunctional (Hutchinson *et al.* in: Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz *et al.*), American Society for Microbiology, Washington D. C. pp. 203-216 (1993)). For the simpler polyketide antibiotics such as actinorhodin (produced by *Streptomyces coelicolor*), the several rounds of two carbon additions are carried out iteratively on PKS enzymes encoded by one set of PKS genes. In contrast, synthesis of the more complicated compounds such as erythromycin and soraphen (see Section E of examples) involves sets of PKS genes organized into modules, with each module carrying out one round of two carbon addition (for review see Hopwood *et al.* in: Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz *et al.*), American Society for Microbiology, Washington D. C.. pp. 267-275 (1993)). The present invention provides the biosynthetic genes of soraphen from *Sorangium* (see Section E of examples). Thus, another embodiment of the present invention relates to an isolated DNA molecule encoding one or more polypeptides for the biosynthesis of soraphen in a heterologous host which molecule can be used to genetically engineer a host organism to express said antipathogenic substance. Other embodiments of the invention are isolated polypeptides required for the biosynthesis of soraphen.

<u>Macrocyclic Lactones</u>

This group of compounds shares the presence of a large lactone ring with various ring substituents. They can be further classified into subgroups, depending on the ring size and other characteristics. The macrolides, for example, contain 12-, 14-, 16-, or 17-membered

lactone rings glycosidically linked to one or more aminosugars and/or deoxysugars. They are inhibitors of protein synthesis, and are particularly effective against gram-positive bacteria. Erythromycin A, a well-studied macrolide produced by *Saccharopolyspora erythraea*, consists of a 14-membered lactone ring linked to two deoxy sugars. Many of the biosynthetic genes have been cloned; all have been located within a 60 kb segment of the *S. erythraea* chromosome. At least 22 closely linked open reading frames have been identified to be likely involved in erythromycin biosynthesis (Donadio *et al.*, in: Industrial Microorganisms: Basic and Applied Molecular Genetics, (ed.: Baltz *et al.*), American Society for Microbiology, Washington D. C.. pp 257-265 (1993)).

Quinones

Quinones are aromatic compounds with two carbonyl groups on a fully unsaturated ring. The compounds can be broadly classified into subgroups according to the number of aromatic rings present, *i.e.*, benzoquinones, napthoquinones, etc. A well studied group is the tetracyclines, which contain a napthacene ring with different substituents. Tetracyclines are protein synthesis inhibitors and are effective against both gram-positive and gram-negative bacteria, as well as rickettsias, mycoplasma, and spirochetes. The aromatic rings in the tetracyclines are derived from polyketide molecules. Genes involved in the biosynthesis of oxytetracycline (produced by *Streptomyces rimosus*) have been cloned and expressed in *Streptomyces lividans* (Binnie *et al.* J. Bacteriol. 171: 887-895 (1989)). The PKS genes share homology with those for actinorhodin and therefore encode type II (monofunctional) PKS proteins (Hopewood & Sherman, Ann. Rev. Genet. 24: 37-66 (1990)).

Other Types of APS

Several other types of APSs have been identified. One of these is the antibiotic 2-hexyl-5-propyl-resorcinol which is produced by certain strains of *Pseudomonas*. It was first isolated from the *Pseudomonas* strain B-9004 (Kanda *et al.* J. Antibiot. 28: 935-942 (1975)) and is a dialkyl-substituted derivative of 1,3-dihydroxybenzene. It has been shown to have antipathogenic activity against Gram-positive bacteria (in particular *Clavibacter* sp.), mycobacteria, and fungi.

Another type of APS are the methoxyacrylates, such as strobilurin B. Strobilurin B is produced by Basidiomycetes and has a broad spectrum of fungicidal activity (Anke, T. *et*

al., Journal of Antibiotics (Tokyo) 30: 806-810 (1977). In particular, strobilurin B is produced by the fungus *Bolinia lutea*. Strobilurin B appears to have antifungal activity as a result of its ability to inhibit cytochrome b dependent electron transport thereby inhibiting respiration (Becker, W. *et al., FEBS Letters 132:* 329-333 (1981).

Most antibiotics have been isolated from bacteria, actinomycetes, and fungi. Their role in the biology of the host organism is often unknown, but many have been used with great success, both in medicine and agriculture, for the control of microbial pathogens. Antibiotics which have been used in agriculture are: blasticidin S and kasugamycin for the control of rice blast (*Pyricularia oryzae*), validamycin for the control of *Rhizoctonia solani*, prumycin for the control of *Botrytis* and *Sclerotinia* species, and mildiomycin for the control of mildew.

To date, the use of antibiotics in plant protection has involved the production of the compounds through chemical synthesis or fermentation and application to seeds, plant parts, or soil. This invention describes the identification and isolation of the biosynthetic genes of a number of anti-phytopathogenic substances and further describes the use of these genes to create transgenic plants with enhanced disease resistance characteristics and also the creation of improved biocontrol strains by expression of the isolated genes in organisms which colonize host plants or the rhizosphere. Furthermore, the availability of such genes provides methods for the production of APSs for isolation and application in antipathogenic formulations.

Methods for Cloning Genes for Antipathogenic Substances

Genes encoding antibiotic biosynthetic genes can be cloned using a variety of techniques according to the invention. The simplest procedure for the cloning of APS genes requires the cloning of genomic DNA from an organism identified as producing an APS, and the transfer of the cloned DNA on a suitable plasmid or vector to a host organism which does not produce the APS, followed by the identification of transformed host colonies to which the APS-producing ability has been conferred. Using a technique such as λ::Tn5 transposon mutagenesis (de Bruijn & Lupski, Gene 27: 131-149 (1984)), the exact region of the transforming APS-conferring DNA can be more precisely defined. Alternatively or additionally, the transforming APS-conferring DNA can be cleaved into smaller fragments

and the smallest which maintains the APS-conferring ability further characterized. Whereas the host organism lacking the ability to produce the APS may be a different species to the organism from which the APS derives, a variation of this technique involves the transformation of host DNA into the same host which has had its APS-producing ability disrupted by mutagenesis. In this method, an APS-producing organism is mutated and non-APS producing mutants isolated, and these are complemented by cloned genomic DNA from the APS producing parent strain. A further example of a standard technique used to clone genes required for APS biosynthesis is the use of transposon mutagenesis to generate mutants of an APS-producing organism which, after mutagenesis, fail to produce the APS. Thus, the region of the host genome responsible for APS production is tagged by the transposon and can be easily recovered and used as a probe to isolate the native genes from the parent strain. APS biosynthetic genes which are required for the synthesis of APSs and which are similar to known APS compounds may be clonable by virtue of their sequence homology to the biosynthetic genes of the known compounds. Techniques suitable for cloning by homology include standard library screening by DNA hybridization.

This invention also describes a novel technique for the isolation of APS biosynthetic genes which may be used to clone the genes for any APS, and is particularly useful for the cloning of APS biosynthetic genes which may be recalcitrant to cloning using any of the above techniques. One reason why such recalcitrance to cloning may exist is that the standard techniques described above (except for cloning by homology) may preferentially lead to the isolation of regulators of APS biosynthesis. Once such a regulator has been identified, however, it can be used using this novel method to isolate the biosynthetic genes under the control of the cloned regulator. In this method, a library of transposon insertion mutants is created in a strain of microorganism which lacks the regulator or has had the regulator gene disabled by conventional gene disruption techniques. The insertion transposon used carries a promoter-less reporter gene (*e.g. lacZ*). Once the insertion library has been made, a functional copy of the regulator gene is transferred to the library of cells (*e.g.* by conjugation or electroporation) and the plated cells are selected for expression of the reporter gene. Cells are assayed before and after transfer of the regulator gene. Colonies which express the reporter gene only in the presence of the regulator gene are insertions adjacent to the promoter of genes regulated by the regulator. Assuming the regulator is specific in its regulation for APS-biosynthetic genes, then the genes tagged by this

procedure will be APS-biosynthetic genes. In a preferred embodiment, the cloned regulator gene is the *gafA* gene described in PCT application WO 94/01561 which regulates the expression of the biosynthetic genes for pyrrolnitrin. Thus, this method is a preferred method for the cloning of the biosynthetic genes for pyrrolnitrin.

An alternative method for identifying and isolating a gene from a microorganism required for the biosynthesis of an antipathogenic substance (APS), wherein the expression of said gene is under the control of a regulator of the biosynthesis of said APS, comprises

(a) cloning a library of genetic fragments from said microorganism into a vector adjacent to a promoterless reporter gene in a vector such that expression of said reporter gene can occur only if promoter function is provided by the cloned fragment;
(b) transforming the vectors generated from step (a) into a suitable host;
(c) identifying those transformants from step (b) which express said reporter gene only in the presence of said regulator; and
(d) identifying and isolating the DNA fragment operably linked to the genetic fragment from said microorganism present in the transformants identified in step (c);

wherein the DNA fragment isolated and identified in step (d) encodes one or more polypeptides required for the biosynthesis of said APS.

In order for the cloned APS genes to be of use in transgenic expression, it is important that all the genes required for synthesis from a particular metabolite be identified and cloned. Using combinations of, or all the techniques described above, this is possible for any known APS. As most APS biosynthetic genes are clustered together in microorganisms, usually encoded by a single operon, the identification of all the genes will be possible from the identification of a single locus in an APS-producing microorganism. In addition, as regulators of APS biosynthetic genes are believed to regulate the whole pathway, then the cloning of the biosynthetic genes via their regulators is a particularly attractive method of cloning these genes. In many cases the regulator will control transcription of the single entire operon, thus facilitating the cloning of genes using this strategy.

Using the methods described in this application, biosynthetic genes for any APS can be cloned from a microorganism. Expression vectors comprising isolated DNA molecules encoding one or more polypeptides for the biosynthesis of an antipathogenic substance

such as pyrrolnitrin and soraphen can be used to transform a heterolgous host. Suitable heterologous hosts are bacteria, fungi, yeast and plants. In a preferred embodiment of the invention the transformed hosts will be able to synthesize an antipathogenic substance not naturally occuring in said host. The host can then be grown under conditions which allow production of said antipathogenic sequence, which can be thus be collected from the host. Using the methods of gene manipulation and transgenic plant production described in this specification, the cloned APS biosynthetic genes can be modified and expressed in transgenic plants. Suitable APS biosynthetic genes include those described at the beginning of this section, *viz.* aminoglycosides and other carbohydrate containing antibiotics (*e.g.* streptomycin), peptide antibiotics (both non-ribosomally and ribosomally synthesized types), nucleoside derivatives and other heterocyclic antibiotics containing nitrogen and/or oxygen (*e.g.* polyoxins, nikkomycins, phenazines, and pyrrolnitrin), polyketides, macrocyclic lactones and quinones (*e.g.* soraphen, erythromycin and tetracycline). Expression in transgenic plants will be under the control of an appropriate promoter and involves appropriate cellular targeting considering the likely precursors required for the particular APS under consideration. Whereas the invention is intended to include the expression in transgenic plants of any APS gene isolatable by the procedures described in this specification, those which are particularly preferred include pyrrolnitrin, soraphen, phenazine, and the peptide antibiotics gramicidin and epidermin. The cloned biosynthetic genes can also be expressed in soil-borne or plant colonizing organisms for the purpose of conferring and enhancing biocontrol efficacy in these organisms. Particularly preferred APS genes for this purpose are those which encode pyrrolnitrin, soraphen, phenazine, and the peptide antibiotics.

Production of Antipathogenic Substances in Heterologous Microbial Hosts

Cloned APS genes can be expressed in heterologous bacterial or fungal hosts to enable the production of the APS with greater efficiency than might be possible from native hosts. Techniques for these genetic manipulations are specific for the different available hosts and are known in the art. For example, the expression vectors pKK223-3 and pKK223-2 can be used to express heterologous genes in *E. coli,* either in transcriptional or translational fusion, behind the *tac or trc* promoter. For the expression of operons encoding multiple ORFs, the simplest procedure is to insert the operon into a vector such as pKK223-3 in transcriptional fusion, allowing the cognate ribosome binding site of the heterologous genes

to be used. Techniques for overexpression in gram-positive species such as *Bacillus* are also known in the art and can be used in the context of this invention (Quax *et al. In.:* Industrial Microorganisms: Basic and Applied Molecular Genetics, *Eds.* Baltz *et al.*, American Society for Microbiology, Washington (1993)). Alternate systems for overexpression rely on yeast vectors and include the use of *Pichia, Saccharomyces* and *Kluyveromyces* (Sreekrishna, *In*: Industrial microorganisms: basic and applied molecular genetics, Baltz, Hegeman, and Skatrud *eds.*, American Society for Microbiology, Washington (1993); Dequin & Barre, Biotechnology 12:173-177 (1994); van den Berg *et al.*, Biotechnology 8:135-139 (1990)).

Cloned APS genes can also be expressed in heterologous bacterial and fungal hosts with the aim of increasing the efficacy of biocontrol strains of such bacterial and fungal hosts. Thus, a method for protecting plants against phytopathogens is to treat said plant with a biocontrol agent transformed with one or more vectors collectively capable of expressing all of the polypeptides necessary to produce an anti-pathogenic substance in amounts which inhibit said phythopathogen. Microorganisms which are suitable for the heterologous overexpression of APS genes are all microorganisms which are capable of colonizing plants or the rhizosphere. As such they will be brought into contact with phytopathogenic fungi, bacteria and nematodes causing an inhibition of their growth. These include gram-negative microorganisms such as *Pseudomonas, Enterobacter* and *Serratia*, the gram-positive microorganism *Bacillus* and the fungi *Trichoderma* and *Gliocladium*. Particularly preferred heterologous hosts are *Pseudomonas fluorescens, Pseudomonas putida, Pseudomonas cepacia, Pseudomonas aureofaciens, Pseudomonas aurantiaca, Enterobacter cloacae, Serratia marscesens, Bacillus subtilis, Bacillus cereus, Trichoderma viride, Trichoderma harzianum* and *Gliocladium virens.* In preferred embodiments of the invention the biosynthetic genes for pyrrolnitrin, soraphen, phenazine, and/or peptide antibiotics are transferred to the particularly preferred heterologous hosts listed above. In a particularly preferred embodiment, the biosynthetic genes for phenazine and/or soraphen are transferred to and expressed in *Pseudomonas fluorescens* strain CGA267356 (described in the published application EP 0 472 494) which has biocontrol utility due to its production of pyrrolnitrin (but not phenazine). In another preferred embodiment, the biosynthetic genes for pyrrolnitrin and/or soraphen are transferred to *Pseudomonas aureofaciens* strain 30-84 which has biocontrol characteristics due to its production of phenazine. Expression in heterologous biocontrol strains requires the selection of vectors appropriate for replication in

the chosen host and a suitable choice of promoter. Techniques are well known in the art for expression in gram-negative and gram-positive bacteria and fungi and are described elsewhere in this specification.

Expression of Genes for Anti-phytopathogenic Substances in Plants

A method for protecting plants against phytopathogens is to transform said plant with one or more vectors collectively capable of expressing all of the polypeptides necessary to produce an anti-pathogenic substance in said plant in amounts which inhibit said phythopathogen. The APS biosynthetic genes of this invention when expressed in transgenic plants cause the biosynthesis of the selected APS in the transgenic plants. In this way transgenic plants with enhanced resistance to phytopathogenic fungi, bacteria and nematodes are generated. For their expression in transgenic plants, the APS genes and adjacent sequences may require modification and optimization.

Although in many cases genes from microbial organisms can be expressed in plants at high levels without modification, low expression in transgenic plants may result from APS genes having codons which are not preferred in plants. It is known in the art that all organisms have specific preferences for codon usage, and the APS gene codons can be changed to conform with plant preferences, while maintaining the amino acids encoded. Furthermore, high expression in plants is best achieved from coding sequences which have at least 35% GC content, and preferably more than 45%. Microbial genes which have low GC contents may express poorly in plants due to the existence of ATTTA motifs which may destabilize messages, and AATAAA motifs which may cause inappropriate polyadenylation. In addition, potential APS biosynthetic genes can be screened for the existence of illegitimate splice sites which may cause message truncation. All changes required to be made within the APS coding sequence such as those described above can be made using well known techniques of site directed mutagenesis, PCR, and synthetic gene construction using the methods described in the published patent applications EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol), and WO 93/07278 (to Ciba-Geigy). The preferred APS biosynthetic genes may be unmodified genes, should these be expressed at high levels in target transgenic plant species, or alternatively may be genes modified by the removal of destabilization and inappropriate polyadenylation motifs and illegitimate splice sites, and further modified by the incorporation of plant preferred codons, and further with a GC content preferred for expression in plants. Although preferred gene sequences may be

adequately expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.* Nucl. Acids Res. 17: 477-498 (1989)).

For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. The sequences cognate to the selected APS genes may initiate translation efficiently in plants, or alternatively may do so inefficiently. In the case that they do so inefficiently, they can be modified by the inclusion of sequences known to be effective in plants. Joshi has suggested an appropriate consensus for plants (NAR 15: 6643-6653 (1987) ; SEQ ID NO:8)) and Clontech suggests a further consensus translation initiator (1993/1994 catalog, page 210; SEQ ID NO:7). These consensuses are suitable for use with the APS biosynthetic genes of this invention. The sequences are incorporated into the APS gene construction, up to and including the ATG (whilst leaving the second amino acid of the APS gene unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

Expression of APS genes in transgenic plants is behind a promoter shown to be functional in plants. The choice of promoter will vary depending on the temporal and spatial requirements for expression, and also depending on the target species. For the protection of plants against foliar pathogens, expression in leaves is preferred; for the protection of plants against ear pathogens, expression in inflorescences (*e.g.* spikes, panicles, cobs *etc.*) is preferred; for protection of plants against root pathogens, expression in roots is preferred; for protection of seedlings against soil-borne pathogens, expression in roots and/or seedlings is preferred. In many cases, however, expression against more than one type of phytopathogen will be sought, and thus expression in multiple tissues will be desirable. Although many promoters from dicotyledons have been shown to be operational in monocotyledons and *vice versa*, ideally dicotyledonous promoters are selected for expression in dicotyledons, and monocotyledonous promoters for expression in monocotyledons. However, there is no restriction to the provenance of selected promoters; it is sufficient that they are operational in driving the expression of the APS biosynthetic genes. In some cases, expression of APSs in plants may provide protection against insect pests. Transgenic expression of the biosynthetic genes for the APS beauvericin (isolated

from *Beauveria bassiana*) may, for example provide protection against insect pests of crop plants.

Preferred promoters which are expressed constitutively include the CaMV 35S and 19S promoters, and promoters from genes encoding actin or ubiquitin. Further preferred constitutive promoters are those from the 12(4-28), CP21, CP24, CP38, and CP29 genes whose cDNAs are provided by this invention.

The APS genes of this invention can also be expressed under the regulation of promoters which are chemically regulated. This enables the APS to be synthesized only when the crop plants are treated with the inducing chemicals, and APS biosynthesis subsequently declines. Preferred technology for chemical induction of gene expression is detailed in the published European patent application EP 0 332 104 (to Ciba-Geigy) herein incorporated by reference. A preferred promoter for chemical induction is the tobacco PR-1a promoter.

A preferred category of promoters is that which is wound inducible. Numerous promoters have been described which are expressed at wound sites and also at the sites of phytopathogen infection. These are suitable for the expression of APS genes because APS biosynthesis is turned on by phytopathogen infection and thus the APS only accumulates when infection occurs. Ideally, such a promoter should only be active locally at the sites of infection, and in this way APS only accumulates in cells which need to synthesize the APS to kill the invading phytopathogen. Preferred promoters of this kind include those described by Stanford *et al.* Mol. Gen. Genet. 215: 200-208 (1989), Xu *et al.* Plant Molec. Biol. 22: 573-588 (1993), Logemann *et al.* Plant Cell 1: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22: 783-792 (1993), Firek *et al.* Plant Molec. Biol. 22: 129-142 (1993), and Warner *et al.* Plant J. 3: 191-201 (1993).

Preferred tissue specific expression patterns include green tissue specific, root specific, stem specific, and flower specific. Promoters suitable for expression in green tissue include many which regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. A preferred promoter is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, Plant Molec. Biol. 12: 579-589 (1989)). A preferred promoter for root specific expression is that

described by de Framond (FEBS 290: 103-106 (1991); EP 0 452 269 to Ciba-Geigy) and a further preferred root-specific promoter is that from the T-1 gene provided by this invention. A preferred stem specific promoter is that described in patent application WO 93/07278 (to Ciba-Geigy) and which drives expression of the maize *trpA* gene.

Preferred embodiments of the invention are transgenic plants expressing APS biosynthetic genes in a root-specific fashion. In an especially preferred embodiment of the invention the biosynthetic genes for pyrrolnitrin are expressed behind a root specific promoter to protect transgenic plants against the phytopathogen *Rhizoctonia*. In another especially preferred embodiment of the invention the biosynthetic genes for phenazine are expressed behind a root specific promoter to protect transgenic plants against the phytopathogen *Gaeumannomyces graminis*. Further preferred embodiments are transgenic plants expressing APS biosynthetic genes in a wound-inducible or pathogen infection-inducible manner. For example, a further especially preferred embodiment involves the expression of the biosynthetic genes for soraphen behind a wound-inducible or pathogen-inducible promoter for the control of foliar pathogens.

In addition to the selection of a suitable promoter, constructions for APS expression in plants require an appropriate transcription terminator to be attached downstream of the heterologous APS gene. Several such terminators are available and known in the art (*e.g.* *tm1* from CaMV, E9 from *rbcS*). Any available terminator known to function in plants can be used in the context of this invention.

Numerous other sequences can be incorporated into expression cassettes for APS genes. These include sequences which have been shown to enhance expression such as intron sequences (*e.g.* from *Adh1* and *bronze1*) and viral leader sequences (*e.g.* from TMV, MCMV and AMV).

The overproduction of APSs in plants requires that the APS biosynthetic gene encoding the first step in the pathway will have access to the pathway substrate. For each individual APS and pathway involved, this substrate will likely differ, and so too may its cellular localization in the plant. In many cases the substrate may be localized in the cytosol, whereas in other cases it may be localized in some subcellular organelle. As much biosynthetic activity in the

plant occurs in the chloroplast, often the substrate may be localized to the chloroplast and consequently the APS biosynthetic gene products for such a pathway are best targeted to the appropriate organelle (*e.g.* the chloroplast). Subcellular localization of transgene encoded enzymes can be undertaken using techniques well known in the art. Typically, the DNA encoding the target peptide from a known organelle-targeted gene product is manipulated and fused upstream of the required APS gene/s. Many such target sequences are known for the chloroplast and their functioning in heterologous constructions has been shown. In a preferred embodiment of this invention the genes for pyrrolnitrin biosynthesis are targeted to the chloroplast because the pathway substrate tryptophan is synthesized in the chloroplast.

In some situations, the overexpression of APS genes may deplete the cellular availability of the substrate for a particular pathway and this may have detrimental effects on the cell. In situations such as this it is desirable to increase the amount of substrate available by the overexpression of genes which encode the enzymes for the biosynthesis of the substrate. In the case of tryptophan (the substrate for pyrrolnitrin biosynthesis) this can be achieved by overexpressing the *trpA and trpB* genes as well as anthranilate synthase subunits. Similarly, overexpression of the enzymes for chorismate biosynthesis such as DAHP synthase will be effective in producing the precursor required for phenazine production. A further way of making more substrate available is by the turning off of known pathways which utilize specific substrates (provided this can be done without detrimental side effects). In this manner, the substrate synthesized is channeled towards the biosynthesis of the APS and not towards other compounds.

Vectors suitable for plant transformation are described elsewhere in this specification. For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one T-DNA border sequence are suitable, whereas for direct gene transfer any vector is suitable and linear DNA containing only the construction of interest may be preferred. In the case of direct gene transfer, transformation with a single DNA species or co-transformation can be used (Schocher *et al.* Biotechnology 4: 1093-1096 (1986)). For both direct gene transfer and *Agrobacterium*-mediated transfer, transformation is usually (but not necessarily) undertaken with a selectable marker which may provide resistance to an antibiotic

(kanamycin, hygromycin or methatrexate) or a herbicide (basta). The choice of selectable marker is not, however, critical to the invention.

Synthesis of an APS in a transgenic plant will frequently require the simultaneous overexpression of multiple genes encoding the APS biosynthetic enzymes. This can be achieved by transforming the individual APS biosynthetic genes into different plant lines individually, and then crossing the resultant lines. Selection and maintenance of lines carrying multiple genes is facilitated if each the various transformation constructions utilize different selectable markers. A line in which all the required APS biosynthetic genes have been pyramided will synthesize the APS, whereas other lines will not. This approach may be suitable for hybrid crops such as maize in which the final hybrid is necessarily a cross between two parents. The maintenance of different inbred lines with different APS genes may also be advantageous in situations where a particular APS pathway may lead to multiple APS products, each of which has a utility. By utilizing different lines carrying different alternative genes for later steps in the pathway to make a hybrid cross with lines carrying all the remaining required genes it is possible to generate different hybrids carrying different selected APSs which may have different utilities.

Alternate methods of producing plant lines carrying multiple genes include the retransformation of existing lines already transformed with an APS gene or APS genes (and selection with a different marker), and also the use of single transformation vectors which carry multiple APS genes, each under appropriate regulatory control (*i.e.* promoter, terminator *etc.*). Given the ease of DNA construction, the manipulation of cloning vectors to carry multiple APS genes is a preferred method.

Before plant propagation material (fruit, tuber, grains, seed) and expecially before seed is sold as a commerical product, it is customarily treated with a protectant coating comprising herbicides, insecticides, fungicides, bactericides, nematicides, molluscicides or mixtures of several of these compounds. If desired these compounds are formulated together with further carriers, surfactants or application-promoting adjuvants customarily employed in the art of formulation to provide protection against damage caused by bacterial, fungal or animal pests.

In order to treat the seed, the protectant coating may be applied to the seeds either by impregnating the tubers or grains with a liquid formulation or by coating them with a

combined wet or dry formulation. In special cases other methods of application to plants are possible such as treatment directed at the buds or the fruit.
A plant seed according to the invention comprises a DNA sequence encoding for the production of an antipathogenic substance and may be treated with a seed protectant coating comprising a seed treatment compound such as captan, carboxin, thiram (TMTD®), methalaxyl (Apron®), pirimiphos-methyl (Actellic®) and others that are commonly used in seed treatment. It is thus a further object of the present invention to provide plant propagation material and especially seed encoding for the production of an antipathogenic substance, which material is treated with a seed protectant coating customarily used in seed treatment.

Production of Antipathogenic Substances in Heterologous Hosts

The present invention also provides methods for obtaining APSs. These APSs may be effective in the inhibition of growth of microbes, particularly phytopathogenic microbes. The APSs can be produced in large quantities from organisms in which the APS genes have been overexpressed, and suitable organisms for this include gram-negative and gram-positive bacteria and yeast, as well as plants. For the purposes of APS production, the significant criteria in the choice of host organism are its ease of manipulation, rapidity of growth (*i.e.* fermentation in the case of microorganisms), and its lack of susceptibility to the APS being overproduced. In a preferred embodiment of the invention enhanced amounts of an antipathogenic substance are synthesized in a host, in which the antipathogenic substance naturally occurs, wherein said host is transformed with one or more DNA molecules collectively encoding the complete set of polypeptides required to synthesize said antipathogenic substance. These methods of APS production have significant advantages over the chemical synthesis technology usually used in the preparation of APSs such as antibiotics. These advantages are the cheaper cost of production, and the ability to synthesize compounds of a preferred biological enantiomer, as opposed to the racemic mixtures inevitably generated by organic synthesis. The ability to produce stereochemically appropriate compounds is particularly important for molecules with many chirally active carbon atoms. APSs produced by heterologous hosts can be used in medical (*i.e.* control of pathogens and/or infectious disease) as well as agricultural applications.

Formulation of Antipathogenic Compositions

The present invention further embraces the preparation of antifungal compositions in which the active ingredient is the antibiotic substance produced by the recombinant biocontrol agent of the present invention or alternatively a suspension or concentrate of the microorganism. The active ingredient is homogeneously mixed with one or more compounds or groups of compounds described herein. The present invention also relates to methods of protecting plants against a phytopathogen, which comprise application of the active ingredient, or antifungal compositions containing the active ingredient, to plants in amounts which inhibit said phytopathogen.

The active ingredients of the present invention are normally applied in the form of compositions and can be applied to the crop area or plant to be treated, simultaneously or in succession, with further compounds. These compounds can be both fertilizers or micronutrient donors or other preparations that influence plant growth. They can also be selective herbicides, insecticides, fungicides, bactericides, nematicides, mollusicides or mixtures of several of these preparations, if desired together with further carriers, surfactants or application-promoting adjuvants customarily employed in the art of formulation. Suitable carriers and adjuvants can be solid or liquid and correspond to the substances ordinarily employed in formulation technology, e.g. natural or regenerated mineral substances, solvents, dispersants, wetting agents, tackifiers, binders or fertilizers.

A preferred method of applying active ingredients of the present invention or an agrochemical composition which contains at least one of the active ingredients is leaf application. The number of applications and the rate of application depend on the intensity of infestation by the corresponding phytopathogen (type of fungus). However, the active ingredients can also penetrate the plant through the roots via the soil (systemic action) by impregnating the locus of the plant with a liquid composition, or by applying the compounds in solid form to the soil, e.g. in granular form (soil application). The active ingredients may also be applied to seeds (coating) by impregnating the seeds either with a liquid formulation containing active ingredients, or coating them with a solid formulation. In special cases, further types of application are also possible, for example, selective treatment of the plant stems or buds.

The active ingredients are used in unmodified form or, preferably, together with the adjuvants conventionally employed in the art of formulation, and are therefore formulated in known manner to emulsifiable concentrates, coatable pastes, directly sprayable or dilutable solutions, dilute emulsions, wettable powders, soluble powders, dusts, granulates, and also encapsulations, for example, in polymer substances. Like the nature of the compositions, the methods of application, such as spraying, atomizing, dusting, scattering or pouring, are chosen in accordance with the intended objectives and the prevailing circumstances. Advantageous rates of application are normally from 50 g to 5 kg of active ingredient (a.i.) per hectare, preferably from 100 g to 2 kg a.i./ha, most preferably from 200 g to 500 g a.i./ha.

The formulations, compositions or preparations containing the active ingredients and, where appropriate, a solid or liquid adjuvant, are prepared in known manner, for example by homogeneously mixing and/or grinding the active ingredients with extenders, for example solvents, solid carriers and, where appropriate, surface-active compounds (surfactants).

Suitable solvents include aromatic hydrocarbons, preferably the fractions having 8 to 12 carbon atoms, for example, xylene mixtures or substituted naphthalenes, phthalates such as dibutyl phthalate or dioctyl phthalate, aliphatic hydrocarbons such as cyclohexane or paraffins, alcohols and glycols and their ethers and esters, such as ethanol, ethylene glycol monomethyl or monoethyl ether, ketones such as cyclohexanone, strongly polar solvents such as N-methyl-2-pyrrolidone, dimethyl sulfoxide or dimethyl formamide, as well as epoxidized vegetable oils such as epoxidized coconut oil or soybean oil; or water.

The solid carriers used e.g. for dusts and dispersible powders, are normally natural mineral fillers such as calcite, talcum, kaolin, montmorillonite or attapulgite. In order to improve the physical properties it is also possible to add highly dispersed silicic acid or highly dispersed absorbent polymers. Suitable granulated adsorptive carriers are porous types, for example pumice, broken brick, sepiolite or bentonite; and suitable nonsorbent carriers are materials such as calcite or sand. In addition, a great number of pregranulated materials of inorganic or organic nature can be used, e.g. especially dolomite or pulverized plant residues.

Depending on the nature of the active ingredient to be used in the formulation, suitable surface-active compounds are nonionic, cationic and/or anionic surfactants having good emulsifying, dispersing and wetting properties. The term "surfactants" will also be understood as comprising mixtures of surfactants.

Suitable anionic surfactants can be both water-soluble soaps and water-soluble synthetic surface-active compounds.

Suitable soaps are the alkali metal salts, alkaline earth metal salts or unsubstituted or substituted ammonium salts of higher fatty acids (chains of 10 to 22 carbon atoms), for example the sodium or potassium salts of oleic or stearic acid, or of natural fatty acid mixtures which can be obtained for example from coconut oil or tallow oil. The fatty acid methyltaurin salts may also be used.

More frequently, however, so-called synthetic surfactants are used, especially fatty sulfonates, fatty sulfates, sulfonated benzimidazole derivatives or alkylarylsulfonates.

The fatty sulfonates or sulfates are usually in the form of alkali metal salts, alkaline earth metal salts or unsubstituted or substituted ammoniums salts and have a 8 to 22 carbon alkyl radical which also includes the alkyl moiety of alkyl radicals, for example, the sodium or calcium salt of lignonsulfonic acid, of dodecylsulfate or of a mixture of fatty alcohol sulfates obtained from natural fatty acids. These compounds also comprise the salts of sulfuric acid esters and sulfonic acids of fatty alcohol/ethylene oxide adducts. The sulfonated benzimidazole derivatives preferably contain 2 sulfonic acid groups and one fatty acid radical containing 8 to 22 carbon atoms. Examples of alkylarylsulfonates are the sodium, calcium or triethanolamine salts of dodecylbenzenesulfonic acid, dibutylnapthalenesulfonic acid, or of a naphthalenesulfonic acid/formaldehyde condensation product. Also suitable are corresponding phosphates, e.g. salts of the phosphoric acid ester of an adduct of p-nonylphenol with 4 to 14 moles of ethylene oxide.

Non-ionic surfactants are preferably polyglycol ether derivatives of aliphatic or cycloaliphatic alcohols, or saturated or unsaturated fatty acids and alkylphenols, said derivatives

containing 3 to 30 glycol ether groups and 8 to 20 carbon atoms in the (aliphatic) hydrocarbon moiety and 6 to 18 carbon atoms in the alkyl moiety of the alkylphenols.

Further suitable non-ionic surfactants are the water-soluble adducts of polyethylene oxide with polypropylene glycol, ethylenediamine propylene glycol and alkylpolypropylene glycol containing 1 to 10 carbon atoms in the alkyl chain, which adducts contain 20 to 250 ethylene glycol ether groups and 10 to 100 propylene glycol ether groups. These compounds usually contain 1 to 5 ethylene glycol units per propylene glycol unit.

Representative examples of non-ionic surfactants are nonylphenolpolyethoxyethanols, castor oil polyglycol ethers, polypropylene/polyethylene oxide adducts, tributylphenoxypolyethoxyethanol, polyethylene glycol and octylphenoxyethoxyethanol. Fatty acid esters of polyoxyethylene sorbitan and polyoxyethylene sorbitan trioleate are also suitable non-ionic surfactants.

Cationic surfactants are preferably quaternary ammonium salts which have, as N-substituent, at least one C8-C22 alkyl radical and, as further substituents, lower unsubstituted or halogenated alkyl, benzyl or lower hydroxyalkyl radicals. The salts are preferably in the form of halides, methylsulfates or ethylsulfates, e.g. stearyltrimethylammonium chloride or benzyldi(2-chloroethyl)ethylammonium bromide.

The surfactants customarily employed in the art of formulation are described, for example, in "McCutcheon's Detergents and Emulsifiers Annual," MC Publishing Corp. Ringwood, New Jersey, 1979, and Sisely and Wood, "Encyclopedia of Surface Active Agents," Chemical Publishing Co., Inc. New York, 1980.

The agrochemical compositions usually contain from about 0.1 to about 99 %, preferably about 0.1 to about 95 %, and most preferably from about 3 to about 90 % of the active ingredient, from about 1 to about 99.9 %, preferably from abut 1 to about 99 %, and most preferably from about 5 to about 95 % of a solid or liquid adjuvant, and from about 0 to about 25 %, preferably about 0.1 to about 25 %, and most preferably from about 0.1 to about 20 % of a surfactant.

Whereas commercial products are preferably formulated as concentrates, the end user will normally employ dilute formulations.

EXAMPLES

The following examples serve as further description of the invention and methods for practicing the invention. They are not intended as being limiting, rather as providing guidelines on how the invention may be practiced.

A. Identification of Microorganisms which Produce Antipathogenic Substances

Microorganisms can be isolated from many sources and screened for their ability to inhibit fungal or bacterial growth *in vitro*. Typically the microorganisms are diluted and plated on medium onto or into which fungal spores or mycelial fragments, or bacteria have been or are to be introduced. Thus, zones of clearing around a newly isolated bacterial colony are indicative of antipathogenic activity.

Example 1: **Isolation of Microorganisms with Anti-*Rhizoctonia* Properties from Soil**

A gram of soil (containing approximately 10^6-10^8 bacteria) is suspended in 10 ml sterile water. After vigorously mixing, the soil particles are allowed to settle. Appropriate dilutions are made and aliquots are plated on nutrient agar plates (or other growth medium as appropriate) to obtain 50-100 colonies per plate. Freshly cultured *Rhizoctonia* mycelia are fragmented by blending and suspensions of fungal fragments are sprayed on to the agar plates after the bacterial colonies have grown to be just visible. Bacterial isolates with antifungal activities can be identified by the fungus-free zones surrounding them upon further incubation of the plates.

The production of bioactive metabolites by such isolates is confirmed by the use of culture filtrates in place of live colonies in the plate assay described above. Such bioassays can also be used for monitoring the purification of the metabolites. Purification may start with an organic solvent extraction step and depending on whether the active principle is extracted into the organic phase or left in the aqueous phase, different chromatographic steps follow.

These chromatographic steps are well known in the art. Ultimately, purity and chemical identity are determined using spectroscopic methods.

B. Cloning Antipathogenic Biosynthetic Genes from Microorganisms

Example 2: **Shotgun Cloning Antipathogenic Biosynthetic Genes from their Native Source**

Related biosynthetic genes are typically located in close proximity to each other in microorganisms and more than one open reading frame is often encoded by a single operon. Consequently, one approach to the cloning of genes which encode enzymes in a single biosynthetic pathway is the transfer of genome fragments from a microorganism containing said pathway to one which does not, with subsequent screening for a phenotype conferred by the pathway.

In the case of biosynthetic genes encoding enzymes leading to the production of an antipathogenic substance (APS), genomic DNA of the antipathogenic substance producing microorganism is isolated, digested with a restriction endonuclease such as *Sau3A*, size fractionated for the isolation of fragments of a selected size (the selected size depends on the vector being used), and fragments of the selected size are cloned into a vector (*e.g.* the *BamHI* site of a cosmid vector) for transfer to *E. coli*. The resulting *E. coli* clones are then screened for those which are producing the antipathogenic substance. Such screens may be based on the direct detection of the antipathogenic substance, such as a biochemical assay.

Alternatively, such screens may be based on the adverse effect associated with the antipathogenic substance upon a target pathogen. In these screens, the clones producing the antipathogenic substance are selected for their ability to kill or retard the growth of the target pathogen. Such an inhibitory activity forms the basis for standard screening assays well known in the art, such as screening for the ability to produce zones of clearing on a bacterial plate impregnated with the target pathogen (eg. spores where the target pathogen is a fungus, cells where the target pathogen is a bacterium). Clones selected for their antipathogenic activity can then be further analyzed to confirm the presence of the

antipathogenic substance using the standard chemical and biochemical techniques appropriate for the particular antipathogenic substance.

Further characterization and identification of the genes encoding the biosynthetic enzymes for the antipathogenic substance is achieved as follows. DNA inserts from positively identified *E. coli* clones are isolated and further digested into smaller fragments. The smaller fragments are then recloned into vectors and reinserted into *E. coli* with subsequent reassaying for the antipathogenic phenotype. Alternatively, positively identified clones can be subjected to λ::Tn5 transposon mutagenesis using techniques well known in the art (*e.g.* de Bruijn & Lupski, Gene 27: 131-149 (1984)). Using this method a number of disruptive transposon insertions are introduced into the DNA shown to confer APS production to enable a delineation of the precise region/s of the DNA which are responsible for APS production. Subsequently, determination of the sequence of the smallest insert found to confer antipathogenic substance production on *E. coli* will reveal the open reading frames required for APS production. These open reading frames can ultimately be disrupted (see below) to confirm their role in the biosynthesis of the antipathogenic substance.

Various host organisms such as *Bacillus* and yeast may be substituted for *E. coli* in the techniques described using suitable cloning vectors known in the art for such host. The choice of host organism has only one limitation; it should not be sensitive to the antipathogenic substance for which the biosynthetic genes are being cloned.

Example 3: **Cloning Biosynthetic Genes for an Antipathogenic Substance using Transposon Mutagenesis**

In many microorganisms which are known to produce antipathogenic substances, transposon mutagenesis is a routine technique used for the generation of insertion mutants. This technique has been used successfully in *Pseudomonas* (e.g. Lam *et al., Plasmid* 13:200-204 (1985)), *Bacillus* (e.g. Youngman *et al., Proc. Natl. Acad. Sci. USA* 80:2305-2309 (1983)), *Staphylococcus* (e.g. Pattee, *J. Bacteriol.* 145:479-488 (1981)), and *Streptomyces* (e.g. Schauer *et al., J. Bacteriol.* 173:5060-5067 (1991)), among others. The main requirement for the technique is the ability to introduce a transposon containing plasmid into the microorganism enabling the transposon to insert itself at a random position in the genome. A large library of insertion mutants is created by introducing a transposon

carrying plasmid into a large number of microorganisms. Introduction of the plasmid into the microorganism can be by any appropriate standard technique such as conjugation, direct gene transfer techniques such as electroporation.

Once a transposon library has been created in the manner described above, the transposon insertion mutants are assayed for production of the APS. Mutants which do not produce the APS would be expected to predominantly occur as the result of transposon insertion into gene sequences required for APS biosynthesis. These mutants are therefore selected for further analysis.

DNA from the selected mutants which is adjacent to the transposon insert is then cloned using standard techniques. For instance, the host DNA adjacent to the transposon insert may be cloned as part of a library of DNA made from the genomic DNA of the selected mutant. This adjacent host DNA is then identified from the library using the transposon as a DNA probe. Alternatively, if the transposon used contains a suitable gene for antibiotic resistance, then the insertion mutant DNA can be digested with a restriction endonuclease which will be predicted not to cleave within this gene sequence or between its sequence and the host insertion point, followed by cloning of the fragments thus generated into a microorganism such as *E. coli* which can then be subjected to selection using the chosen antibiotic.

Sequencing of the DNA beyond the inserted transposon reveals the adjacent host sequences. The adjacent sequences can in turn be used as a hybridization probe to reclone the undisrupted native host DNA using a non-mutant host library. The DNA thus isolated from the non-mutant is characterized and used to complement the APS deficient phenotype of the mutant. DNA which complements may contain either APS biosynthetic genes or genes which regulate all or part of the APS biosynthetic pathway. To be sure isolated sequences encode biosynthetic genes they can be transferred to a heterologous host which does not produce the APS and which is insensitive to the APS (such as *E. coli*). By transferring smaller and smaller pieces of the isolated DNA and the sequencing of the smallest effective piece, the APS genes can be identified. Alternatively, positively identified clones can be subjected to λ::Tn5 transposon mutagenesis using techniques well known in the art (*e.g.* de Bruijn & Lupski, Gene 27: 131-149 (1984)). Using this method a number of

disruptive transoposon insertions are introduced into the DNA shown to confer APS production to enable a delineation of the precise region/s of the DNA which are responsible for APS production. These latter steps are undertaken in a manner analagous to that described in example 1. In order to avoid the possibility of the cloned genes not being expressed in the heterologous host due to the non-functioning of their heterologous promoter, the cloned genes can be transferred to an expression vector where they will be fused to a promoter known to function in the heterologous host. In the case of *E. coli* an example of a suitable expression vector is pKK223 which utilizes the *tac* promoter. Similar suitable expression vectors also exist for other hosts such as yeast and are well known in the art. In general such fusions will be easy to undertake because of the operon-type organization of related genes in microorganisms and the likelihood that the biosynthetic enzymes required for APS biosynthesis will be encoded on a single transcript requiring only a single promoter fusion.

Example 4: **Cloning Antipathogenic Biosynthetic Genes using Mutagenesis and Complementation**

A similar method to that described above involves the use of non-insertion mutagenesis techniques (such as chemical mutagenesis and radiation mutagenesis) together with complementation. The APS producing microorganism is subjected to non-insertion mutagenesis and mutants which lose the ability to produce the APS are selected for further analysis. A gene library is prepared from the parent APS-producing strain. One suitable approach would be the ligation of fragments of 20-30 kb into a vector such as pVK100 (Knauf *et al.* Plasmid 8: 45-54 (1982)) into *E. coli* harboring the tra+ plasmid pRK2013 which would enable the transfer by triparental conjugation back to the selected APS-minus mutant (Ditta *et al.* Proc. Natl. Acad. Sci. USA 77: 7247-7351 (1980)). A further suitable approach would be the transfer back to the mutant of the genes library via electroporation. In each case subsequent selection is for APS production. Selected colonies are further characterized by the retransformation of APS-minus mutant with smaller fragments of the complementing DNA to identify the smallest successfully complementing fragment which is then subjected to sequence analysis. As with example 2, genes isolated by this procedure may be biosynthetic genes or genes which regulate the entire or part of the APS biosynthetic pathway. To be sure that the isolated sequences encode biosynthetic genes they can be transferred to a heterologous host which does not produce the APS and is

insensitive to the APS (such as *E. coli*). These latter steps are undertaken in a manner analagous to that described in example 2.

Example 5: **Cloning Antipathogenic Biosynthetic Genes by Exploiting Regulators which Control the Expression of the Biosynthetic Genes**

A further approach in the cloning of APS biosynthetic genes relies on the use of regulators which control the expression of these biosynthetic genes. A library of transposon insertion mutants is created in a strain of microorganism which lacks the regulator or has had the regulator gene disabled by conventional gene disruption techniques. The insertion transposon used carries a promoter-less reporter gene (*e.g. lacZ*). Once the insertion library has been made, a functional copy of the regulator gene is transferred to the library of cells (*e.g.* by conjugation or electroporation) and the plated cells are selected for expression of the reporter gene. Cells are assayed before and after transfer of the regulator gene. Colonies which express the reporter gene only in the presence of the regulator gene are insertions adjacent to the promoter of genes regulated by the regulator. Assuming the regulator is specific in its regulation for APS-biosynthetic genes, then the genes tagged by this procedure will be APS-biosynthetic genes. These genes can then be cloned and further characterized using the techniques described in example 2.

Example 6: **Cloning Antipathogenic Biosynthetic Genes by Homology**

Standard DNA techniques can be used for the cloning of novel antipathogenic biosynthetic genes by virtue of their homology to known genes. A DNA library of the microorganism of interest is made and then probed with radiolabelled DNA derived from the gene/s for APS biosynthesis from a different organism. The newly isolated genes are characterized and sequenced and introduced into a heterologous microorganism or a mutant APS-minus strain of the native microorganisms to demonstrate their conferral of APS production.

C. Cloning of Pyrrolnitrin Biosynthetic Genes from *Pseudomonas*

Pyrrolnitrin is a phenylpyrole compound produced by various strains of *Pseudomonas fluorescens*. *P. fluorescens* strains which produce pyrrolnitrin are effective biocontrol strains against *Rhizoctonia* and *Pythium* fungal pathogens (WO 94/01561). The biosynthesis of pyrrolnitrin is postulated to start from tryptophan (Chang *et al.* J. Antibiotics 34: 555-566 (1981)).

Example 7: **Use of the *gafA* Regulator Gene for the Isolation of Pyrrolnitrin Biosynthetic Genes from *Pseudomonas***

The gene cluster encoding pyrrolnitrin biosynthetic enzymes was isolated using the basic principle described in example 5 above. The regulator gene used in this isolation procedure was the *gafA* gene from *Pseudomonas fluorescens* and is known to be part of a two-component regulatory system controlling certain biocontrol genes in *Pseudomonas*. The *gafA* gene is described in detail in WO 94/01561 which is hereby incorporated by reference in its entirety. *gafA* is further described in Gaffney *et al.* (Molecular Plant-Microbe Interactions 7: 455-463, 1994, also hereby incorporated in its entirety by reference) where it is referred to as "ORF5". The *gafA* gene has been shown to regulate pyrrolnitrin biosynthesis, chitinase, gelatinase and cyanide production. Strains which lack the *gafA* gene or which express the gene at low levels (and in consequence *gafA*-regulated genes also at low levels) are suitable for use in this isolation technique.

Example 8: **Isolation of Pyrrolnitrin Biosynthesis Genes in *Pseudomonas***

The transfer of the *gafA* gene from MOCG 134 to closely related non-pyrrolnitrin producing wild-type strains of *Pseudomonas fluorescens* results in the ability of these strains to produce pyrrolnitrin. (Gaffney *et al.*, MPMI (1994)); see also Hill *et al.* Applied And Environmental Microbiology 60 78-85 (1994)). This indicates that these closely related strains have the structural genes needed for pyrrolnitrin biosynthesis but are unable to produce the compound without activation from the *gafA* gene. One such closely related strain, MOCG133, was used for the identification of the pyrrolnitrin biosynthesis genes. The transposon TnCIB116 (Lam, New Directions in Biological Control: Alternatives for Suppressing Agricultural Pests and Diseases, pp 767-778, Alan R. Liss, Inc. (1990)) was used to mutagenize MOCG133. This transposon, a Tn5 derivative, encodes kanamycin resistance and contains a promoterless lacZ reporter gene near one end. The transposon was introduced into MOCG133 by conjugation, using the plasmid vector pCIB116 (Lam, New Directions in Biological Control: Alternatives for Suppressing Agricultural Pests and Diseases, pp 767-778, Alan R. Liss, Inc. (1990)) which can be mobilized into MOCG133, but cannot replicate in that organism. Most, if not all, of the kanamycin resistant transconjugants were therefore the result of transposition of TnCIB116 into different sites in the MOCG133 genome. When the transposon integrates into the bacterial chromosome behind an active promoter the lacZ reporter gene is activated. Such gene activation can be

monitored visually by using the substrate X-gal, which releases an insoluble blue product upon cleavage by the lacZ gene product. Kanamycin resistant transconjugants were collected and arrayed on master plates which were then replica plated onto lawns of E coli strain S17-1 (Simon *et al.*, Bio/techonology 1:784-791 (1983)) transformed with a plasmid carrying the wide host range RK2 origin of replication, a gene for tetracycline selection and the *gafA* gene. *E coli* strain S17-1 contains chromosomally integrated tra genes for conjugal transfer of plasmids. Thus, replica plating of insertion transposon mutants onto a lawn of the S17-1/*gafA E. coli* results in the transfer to the insertion transposon mutants of the *gafA*-carrying plasmid and enables the activity of the *lacZ* gene to be assayed in the presence of the *gafA* regulator (expression of the host *gafA* is insufficient to cause *lacZ* expression, and introduction of *gafA* on a multicopy plasmid is more effective). Insertion mutants which had a "blue" phenotype (i.e. lacZ activity) only in the presence of gafA were identified. In these mutants, the transposon had integrated within genes whose expression were regulated by gafA. These mutants (with introduced gafA) were assayed for their ability to produce cyanide, chitinase, and pyrrolnitrin (as described in Gaffney *et al.*, 1994 MPMI, in press) --activities known to be regulated by gafA (Gaffney *et al.*, 1994 MPMI, in press). One mutant did not produce pyrrolnitrin but did produce cyanide and chitinase, indicating that the transposon had inserted in a genetic region involved only in pyrrolnitrin biosynthesis. DNA sequences flanking one end of the transposon were cloned by digesting chromosomal DNA isolated from the selected insertion mutant with *XhoI*, ligating the fragments derived from this digestion into the *XhoI* site of pSP72 (Promega, cat. # P2191) and selecting the *E. coli* transformed with the products of this ligation on kanamycin. The unique *XhoI* site within the transposon cleaves beyond the gene for kanamycin resistance and enabled the flanking region derived from the parent MOCG 133 strain to be concurrently isolated on the same *XhoI* fragment. In fact the *XhoI* site of the flanking sequence was found to be located approximately 1 kb away from the end on the transposon. A subfragment of the cloned *XhoI* fragment derived exclusively from the ~1 kb flanking sequence was then used to isolate the native (*i.e.* non-disrupted) gene region from a cosmid library of strain MOCG 134. The cosmid library was made from partially *Sau3A* digested MOCG 134 DNA, size selected for fragments of between 30 and 40 kb and cloned into the unique *BamHI* site of the cosmid vector pCIB119 which is a derivative of c2XB (Bates & Swift, Gene 26: 137-146 (1983)) and pRK290 (Ditta *et al.* Proc. Natl. Acad. Sci. USA 77: 7247-7351 (1980)). pCIB119 is a double-*cos* site cosmid vector which has the

wide host range RK2 origin of replication and can therefore replicate in *Pseudomonas* as well as *E. coli.* Several clones were isolated from the MOCG 134 cosmid clone library using the ~1 kb flanking sequence as a hybridization probe. Of these one clone was found to restore pyrrolnitrin production to the transposon insertion mutant which had lost its ability to produce pyrrolnitrin. This clone had an insertion of ~32 kb and was designated pCIB169. A viable culture of E.coli DH5α comprising cosmid clone pCIB169 has been deposited with the Agricultural Research Culture Collection (NRRL) at 1815 N. University Street, Peoria, Illinois 61604 U.S.A. on May 20, 1994, under the accession number NRRL B-21256.

Example 9: **Mapping and Tn5 Mutagenesis of pCIB169**

The 32 kb insert of clone pCIB169 was subcloned into pCIB189 in *E coli* HB101, a derivative of pBR322 which contains a unique *NotI* cloning site. A convenient *NotI* site within the 32 kb insert as well as the presence of *NotI* sites flanking the *BamHI* cloning site of the parent cosmid vector pCIB119 allowed the subcloning of fragments of 14 and 18 kb into pCIB189. These clones were both mapped by restriction digestion and figure 1 shows the result of this. λ Tn5 transposon mutagenesis was carried out on both the 14 and 18 kb subclones using techniques well known in the art (*e.g.* de Bruijn & Lupski, Gene 27: 131-149 (1984). λ Tn5 phage conferring kanamycin resistance was used to transfect both the 14 and the 18 kb subclones described above. λ Tn5 transfections were done at a multiplicity of infection of 0.1 with subsequent selection on kanamycin. Following mutagenesis plasmid DNA was prepared and retransformed into *E coli* HB101 with kanamycin selection to enable the isolation of plasmid clones carrying Tn5 insertions. A total of 30 independent Tn5 insertions were mapped along the length of the 32 kb insert (see figure 2). Each of these insertions was crossed into MOCG 134 via double homologous recombination and verified by Southern hybridization using the Tn5 sequence and the pCIB189 vector as hybridization probes to demonstrate the occurrence of double homologous recombination *i.e.* the replacement of the wild-type MOCG 134 gene with the Tn5-insertion gene. Pyrrolnitrin assays were performed on each of the insertions that were crossed into MOCG 134 and a genetic region of approximately 6 kb was identified to be involved in pyrrolnitrin production (see figures 3 and 5). This region was found to be centrally located in pCIB169 and was easily subcloned as an *XbaI/NotI* fragment into pBluescript II KS (Promega). The *XbaI/NotI* subclone was designated pPRN5.9X/N (see figure 4).

Example 10: **Identification of Open Reading Frames In the Cloned Genetic Region**

The genetic region involved in pyrrolnitrin production was subcloned into six fragments for sequencing in the vector pBluescript II KS (see figure 4). These fragments spanned the ~6 kb *XbaI/NotI* fragment described above and extended from the *EcoRI* site on the left side of figure 4 to the rightmost *HindIII* site (see figure 4). The sequence of the inserts of clones pPRN1.77E, pPRN1.01E, pPRN1.24E, pPRN2.18E, pPRN0.8H/N, and pPRN2.7H was determined using the Taq DyeDeoxy Terminator Cycle Sequencing Kit supplied by Applied Biosystems, Inc., Foster City, CA. following the protocol supplied by the manufacturer. Sequencing reactions were run on an Applied Biosystems 373A Automated DNA Sequencer and the raw DNA sequence was assembled and edited using the "INHERIT" software package also from Applied Biosystems, Inc.. A contiguous DNA sequence of 9.7 kb was obtained corresponding to the *EcoRI/HindIII* fragment of Figure 3 and bounded by *EcoRI* site # 2 and *HindIII* site # 2 depicted in figure 4.

DNA sequence analysis was performed on the contiguous 9.7 kb sequence using the GCG software package from Genetics Computer Group, Inc. Madison,WI. The pattern recognition program "FRAMES" was used to search for open reading frames (ORFs) in all six translation frames of the DNA sequence. Four open reading frames were identified using this program and the codon frequency table from ORF2 of the *gafA* gene region which was previously published (*WO 94/05793*; figure 5). These ORFs lie entirely within the ~6 kb *Xba I/NotI* fragment referred to in example 9 (figure 4) and are contained within the sequence disclosed as SEQ ID NO:1. By comparing the codon frequency usage table from MOCG134 DNA sequence of the *gafA* region to these four open reading frames, very few rare codons were used indicating that codon usage was similar in both of these gene regions. This strongly suggested that the four open reading frames were real. At a 3' position to the fourth reading frame numerous ρ-independent stem loop structures were found suggesting a region where transcription could be stopped. It was thus apparent that all four ORFs were translated from a single transcript. Sequence data obtained for the regions beyond the four identified ORFs revealed a fifth open reading frame which was subsequently determined to not be involved in pyrrolnitrin synthesis based on *E. coli* expression studies.

For each open reading frame (ORF) in the pyrrolnitrin gene cluster multiple putative translation start sites were identified by the presence of an in-frame start codon (ATG or GTG) and an upstream ribosome binding site. A complementation approach was used to identify the actual translation start site for each gene. PCR primers were synthesized to amplify segments of each prn gene from upstream of one of the putative ribosome binding sites to downstream of the stop codon (Table 1). The plasmid pPRN18Not (1506 CIP3, Figure 4) was used as the template for PCR reactions. The PCR products were cloned in the vector pRK(KK223-3MCS) which consists of the Ptac promoter and rrs terminator from pKK223-3 (Pharmacia) and pRK290 backbone. Plasmids containing each construct were mobilized into the respective ORF-deletion mutants of MOCG134 as described in example 12 and by triparental matings using the helper plasmid pRK290 in E. coli HB101. Transconjugants were selected by plating on Pseudomonas minimal medium supplemented with 30 mg/l tetracycline. The presence of the plasmids and correct orientations of the inserted PCR product were verified by plasmid DNA preparation, restriction digestion and agarose gel electrophoresis. Pyrrolnitrin production was determined by extraction and TLC assay as in example 11. For each prn gene the shortest clone restoring pyrrolnitrin production (i.e., complementing the ORF deletion) was judged to contain the actual translation initiation site. Thus, the initiation codons were identified as follows: ORF1 - ATG at nucleotide position 423, ORF2 - GTG at nucleotide position 2026, ORF3 - ATG at nucleotide position 3166, and ORF4 - ATG at nucleotide position 4894. The pattern "FRAMES" computer program used to indentify the open reading frames only recognizes ATG start codons. Using the complementation approach describe here it was determined that ORF2 actually starts with a GTG codon at nucleotide position 2039 and is thus longer than the open reading frame identified by the "FRAMES" program.

Table 1: DNA constructs and hosts used to identify translation initiation sites in the pyrrolnitrin gene cluster[a].

Construct	Start of amplified segment	Putative start codon[b]	Stop codon[c]	End of amplified segment	Host strain[d]	Pyrrolnitrin production
ORF1-1	294	357	2039	2056	ORF1D	+
ORF1-2	396	423	2039	2056	ORF1D	+
ORF1-3	438	477	2039	2056	ORF1D	-
ORF2-1	2026	2039	3076	3166	ORF2D	+
ORF2-2	2145	2162	3076	3166	ORF2D	-
ORF2-3	2249	2215	3076	3166	ORF2D	-
ORF3-1	3130	3166	4869	4904	ORF3D	+
ORF3-2	3207	3235	4869	4904	ORF3D	-
ORF3-3	3329	3355	4869	4904	ORF3D	-
ORF4-1	4851	4894	5985	6122	ORF4D	+
ORF4-2	4967	4990	5985	6122	ORF4D	-
ORF4-3	5014	5086	5985	6122	ORF4D	-

[a] All nucleotide position numbers refer to the Sequence of the Pyrrolnitrin Gene Cluster given in SEQ ID No. 1
[b] The first base of the putative start codon
[c] The last base of the stop codon
[d] ORF deletion mutants are described in Example 12

Example 11: **Expression of Pyrrolnitrin Biosynthetic Genes in *E. coli***

To determine if only four genes were needed for pyrrolnitrin production, these genes were transferred into *E. coli* which was then assayed for pyrrolnitrin production. The expression vector pKK223-3 was used to over-express the cloned operon in *E. coli*. (Brosius & Holy, Proc. Natl. Acad. Sci. USA 81: 6929 (1984)). pKK223-3 contains a strong *tac* promoter which, in the appropriate host, is regulated by the *lac* repressor and induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) to the bacterial growth medium. This vector was modified by the addition of further useful restriction sites to the existing multiple cloning site to facilitate the cloning of the ~6 kb *XbaI/NotI* fragment (see example 7 and figure 4) and a

10 kb *XbaI/KpnI* fragment (see figure 4) for expression studies. In each case the cloned fragment was under the control of the *E. coli tac* promoter (with IPTG induction), but was cloned in a transcriptional fusion so that the ribosome binding site used would be that derived from *Pseudomonas*. Each of these clones was transformed into *E. coli* XL1-blue host cells and induced with 2.5 mM IPTG before being assayed for pyrrolnitrin by thin layer chromatography. Cultures were grown for 24 h after IPTG induction in 10 ml L broth at 37 C with rapid shaking, then extracted with an equal volume of ethyl acetate. The organic phase was recovered, allowed to evaporated under vacuum and the residue dissolved in 20 l of methanol. Silica gel thin layer chromatography (TLC) plates were spotted with 10 l of extract and run with toluene as the mobile phase. The plates were allowed to dry and sprayed with van Urk's reagent to visualize. Urk's reagent comprises 1g p-Dimethylaminobenzaldehyde in 50 ml 36% HCL and 50 ml 95% ethanol. Under these conditions pyrrolnitrin appears as a purple spot on the TLC plate. This assay confirmed the presence of pyrrolnitrin in both of the expression constructs. HPLC and mass spectrometry analysis further confirmed the presence of pyrrolnitrin in both of the extracts. HPLC analysis can be undertaken directly after redissolving in methanol (in this case the sample is redissolved in 55 % methanol) using a Hewlett Packard Hypersil ODS column (5 µM) of dimensions 100 x 2.1 mm.. Pyrrolnitrin elutes after about 14 min.

Example 11a: **Construction of strain MOCG134cPrn having pyrrolnitrin biosynthetic genes under a constitutive promoter**

Transcription of the pyrrolnitrin biosynthetic genes is regulated by gafA. Thus, transcription and Pyrrolnitirin production does not reach high levels until late log and stationary growth phase. To increase pyrrolnitrin biosynthesis in earlier growth phases the endogenous promoter was replaced with the strong constitutive *E. coli tac* promoter. The Prn genes were cloned between the *tac* promoter and a strong terminator sequence as described in example 11 above. The resulting synthetic operon was inserted into a genomic clone that had the Prn biosynthetic genes deleted but has homologous sequences both upstream and downstream of the insertion site. This clone was mobilized into strain MOCG134_Prn, a deletion mutant of the genes Prn A-D. The Prn genes under the control of the constitutive *tac* promoter were inserted into the bacterial chromosome via double homologous recombination. The resultant strain MOCG134cPrn was shown to produce Pyrrolnitrin earlier than the wild-type strain.

Pyrrolnitrin production of the wild type strain MOCG134, of strain MOCG134cPrn, and of a strain containing plasmid borne PRN genes under the control of the *tac* promoter (MOCG134pPrn) was assayed at various time points (14, 17, 20, 23 and 26 hours growth). Cultures were inoculated with a 1/10,000 dilution of a stationary phase culture, Pyrrolnitrin was extracted with ethyl acetate, and the amount of Pyrrolnitrin was determined by integrating the peak area of Pyrrolnitrin detected by HPLC at 212 nm. The results shown in Table 3 clearly indicate that strains containing the Prn genes under the control of the *tac* promoter produce Pyrronnitrin much earlier than the wilde type MOCG134 strain. The new strains produce Pyrrolnitrin independent of gafA and are useful as new biocontrol strains.

Table 3: Pyrrolnitrin production of different strains at different time points

	amount Pyrrolnitrin produced (peak area)		
time of growth (hours)	MOCG134	MOCG134cPrn	MOCG134pPrn
14	1250	7100	18300
17	3500	14600	26700
20	9600	16600	32100
23	17500	18900	31000
26	25000	22500	33500

Example 12: **Construction of Pyrrolnitrin Gene Deletion Mutants**

To further demonstrate the involvement of the 4 ORFs in pyrrolnitrin biosynthesis, independent deletions were created in each ORF and transferred back into *Pseudomonas fluorescens* strain MOCG134 by homologous recombination. The plasmids used to generate deletions are depicted in Figure 4 and the positions of the deletions are shown in Figure 6. Each ORF is identified within the sequence disclosed as SEQ ID NO:1.

ORF1 (SEQ ID NO:2):

The plasmid pPRN1.77E was digested with *Mlu1* to liberate a 78 bp fragment internally from ORF1. The remaining 4.66 kb vector-containing fragment was recovered, religated with T4 DNA ligase, and transformed into the *E. coli* host strain DH5α. This new plasmid was linearized with *Mlu1* and the Klenow large fragment of DNA polymerase I was used to create blunt ends (Maniatis *et al.* Molecular Cloning, Cold Spring Harbor Laboroatory

(1982)). The neomycin phosphotransferase II (NPTII) gene cassette from pUC4K (Pharmacia) was ligated into the plasmid by blunt end ligation and the new construct, designated pBS(ORF1Δ), was transformed into DH5α. The construct contained a 78 bp deletion of ORF1 at which position the NPTII gene conferring kanamycin resistance had been inserted. The insert of this plasmid (*i.e.* ORF1 with NPTII insertion) was then excised from the pBluescript II KS vector with *EcoRI*, ligated into the *EcoRI* site of the vector pBR322 and transformed into the *E. coli* host strain HB101. The new plasmid was verified by restriction enzyme digestion and designated pBR322(ORF1Δ).

ORF2 (SEQ ID NO:3):

The plasmids pPRN1.24E and pPRN1.01E containing contiguous *EcoRI* fragments spanning ORF2 were double digested with *EcoRI* and *XhoI*. The 1.09 kb fragment from pPRN1.24E and the 0.69 Kb fragment from pPRN1.01E were recovered and ligated together into the *EcoRI* site of pBR322. The resulting plasmid was transformed into the host strain DH5α and the construct was verified by restriction enzyme digestion and electrophoresis. The plasmid was then linearized with *XhoI*, the NPTII gene cassette from pUC4K was inserted, and the new construct, designated pBR(ORF2Δ), was transformed into HB101. The construct was verified by restriction digestions and agarose gel electrophoresis and contains NPTII within a 472 bp deletion of the ORF2 gene.

ORF3 (SEQ ID NO:4):

The plasmid pPRN2.56Sph was digested with *PstI* to liberate a 350 bp fragment. The remaining 2.22 kb vector-containing fragment was recovered and the NPTII gene cassette from pUC4K was ligated into the *PstI* site. This intermediate plasmid, designated pUC(ORF3Δ), was transformed into DH5α and verified by restriction digestion and agarose gel electrophoresis. The gene deletion construct was excised from pUC with *SphI* and ligated into the SphI site of pBR322. The new plasmid, designated pBR(ORF5Δ), was verified by restriction enzyme digestion and agarose gel electrophoresis. This plasmid contains the NPTII gene within a 350 bp deletion of the ORF3 gene.

ORF4 (SEQ ID NO:5):

The plasmid pPRN2.18E/N was digested with *AatII* to liberate 156 bp fragment. The remaining 2.0 kb vector-containing fragment was recovered, religated, transformed into

DH5α, and verified by restriction enzyme digestion and electrophoresis. The new plasmid was linearized with *AatII* and T4 DNA polymerase was used to create blunt ends. The NPTII gene cassette was ligated into the plasmid by blunt-end ligation and the new construct, designated pBS(ORF4Δ), was transformed into DH5α. The insert was excised from the pBluescript II KS vector with *EcoRI*, ligated into the *EcoRI* site of the vector pBR322 and transformed into the *E. coli* host strain HB101. The identity of the new plasmid, designated pBR(ORF4Δ), was verified by restriction enzyme digestion and agarose gel electrophoresis. This plasmid contains the NPTII gene within a 264 bp deletion of the ORF4 gene.

Km^R Control:

To control for possible effects of the kanamycin resistance marker, the NPTII gene cassette from pUC4K was inserted upstream of the pyrrolnitrin gene region. The plasmid pPRN2.5S (a subclone of pPRN7.2E) was linearized with *PstI* and the NPTII cassette was ligated into the PstI site. This intermediate plasmid was transformed into DH5α and verified by restriction digestions and agarose gel electrophoresis. The gene insertion construct was excised from pUC with *SphI* and ligated into the *SphI* site of pBR322. The new plasmid, designated pBR(2.5SphIKm^R), was verified by restriction enzyme digestion and agarose gel electrophoresis. It contains the NPTII region inserted upstream of the pyrrolnitrin gene region.

Each of the gene deletion constructs was mobilized into MOCG134 by triparental mating using the helper plasmid pRK2013 in *E. coli* HB101. Gene replacement mutants were selected by plating on *Pseudomonas* Minimal Medium (PMM) supplemented with 50 µg/ml kanamycin and counterselected on PMM supplemented with 30 µg/ml tetracycline. Putative perfect replacement mutants were verified by Southern hybridization by probing *EcoRI* digested DNA with pPRN18Not, pBR322 and an NPTII cassette obtained from pUC4K (Pharmacia 1994 catalog no. 27-4958-01). Verification of perfect hybridization was apparent by lack of hybridization to pBR322, hybridization of pPRN18Not to an appropriately size-shifted *EcoRI* fragment (reflecting deletion and insertion of NPTII), hybridization of the NPTII probe to the shifted band, and the disappearance of a band corresponding a deleted fragment.

After verification, deletion mutants were tested for production of pyrrolnitrin, 2-hexyl-5-propyl-resorcinol, cyanide, and chitinase production. A deletion in any one of the ORFs abolished pyrrolnitrin production, but did not affect production of the other substances. The presence of the NPTII gene cassette in the Km^R control had no effect on the production of pyrolnitrin, 2-hexyl-5-propyl-resorcinol, cyanide or chitinase. These experiments demonstrated the requirement of each of the four ORFs for pyrrolnitrin production.

Example 12a: **Cloning of the coding regions for expression in plants**

The coding regions of ORFs 1,2,3, and 4 were designated prnA, prnB, prnC and prnD, respectively. Primers were designed to PCR amplify the coding regions for each prn gene from the start codon to or beyond the stop codon as shown in Table 2. Additionally, the primers were designed to add restriction sites to the ends of the coding regions and in the case of prnB to change the initiation codon for prnB from GTG to ATG. Plasmid pPRN18Not (Figure 4) was used as template for the PCR reactions. The PCR products were cloned into pPEH14 for functional testing. Plasmid pPEH14 is a modification of pRK(KK223-3) which contains a synthetic ribosome binding site 11 to 14 bases upstream of the start codons of the cloned PCR products. The constructs were mobilized into the respective ORF deletion mutants by triparental matings as described earlier. The presence of each plasmid and the correct orientation of the inserted PCR product were confirmed by plasmid DNA extraction, restriction digestion, and agarose gel electrophoresis. Pyrrolnitrin production of the complemented mutants was confirmed as described in example 11.

After the expression of a functional protein by each coding region was verified (i.e., the ability to restore pyrrolnitrin production to an ORF deletion mutant was demonstrated) the clones were sequenced and compared to the sequence of the pyrrolnitrin gene cluster (1506 CIP3). For prnA, prnB and prnC the sequence of the amplified coding regions were identical to the original gene cluster sequences. For prnD there was a single base change at nucleotide position 5605 from G in the original sequence to A in the amplified coding region. This base change results in a change from glycine to serine in the deduced amino acid sequence, but does not affect function of the gene product according to the complementation tests described above.

Table 2: Coding regions of the prn genes[a]

Coding region	Start of amplified segment	Start codon[b]	Stop codon[c]	End of amplified segment
prnA	423	423	2039	2055
prnB	2039	2039	3076	3081
prnC	3166	3166	4869	4075
prnD	4894	4894	5985	5985

[a] All nucleotide position numbers refer to Sequence ID No. 1
[b] The first base of the start codon.
[c] The last base of the codon.

Example 12b: **Expression of prn genes in plants**

The coding regions for each prn gene, described in example 12a above were subcloned into a plant expression cassette consisting of the CaMV 35S promoter and leader and the CaMV 35S terminator flanked by Xba I restriction sites. Each construct comprising promoter, coding region, and terminator was liberated with Xba I, subcloned into the binary transformation vector pCIB200, and then transformed into *Agrobacterium tumifaciens* host strain A136. Tobacco transformation was carried out as described by Horsch et al., Science 227: 1229-1231, 1985). Arabidopsis transformation was carried out as described by Lloyd et al, Science 234:464-466, 1986. Plantlets were selected and regenerated on medium containing 100mg/L kanamycin and 500 mg/L carbenecillin.

Tobacco leaf tissue was harvested from individual plants that were suspected to be transformed. *Arabidopsis* leaf tissue from about 10 independent plants suspected to be transformed was pooled for each gene construct used for transformation. RNA was purified by phenol:chloroform extraction and fractionated by formaldehyde gel electrophoresis before blotting onto nylon membranes. Probes to each coding region were made using the random primed labeling method. Hybridization was carried out in 50% formamide at 42°C as described by Sambrook et al., Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory, 1989.

For each prn gene, transgenic tobacco plants were identified which produced RNA bands hybridizing strongly to the appropriate prn gene probe and showing the size expected for a mRNA transcribed from the relevant prn gene. Similiar bands were also seen in RNA extracted from the pooled samples of *Arabidopsis* tissue. The data demonstrate that mRNAs encoding the enzymes of the pyrrolnitrin biosynthetic pathway accumulate in transgenic plants.

D. Cloning of Resorcinol Biosynthetic Genes from *Pseudomonas*

2-hexyl-5-propyl-resorcinol is a further APS produced by certain strains of *Pseudomonas*. It has been shown to have antipathogenic activity against Gram-positive bacteria (in particular *Clavibacter* spp.), mycobacteria, and fungi.

Example 13: **Isolation of Genes Encoding Resorcinol**

Two transposon-insertion mutants have been isolated which lack the ability to produce the antipathogenic substance 2-hexyl-5-propyl-resorcinol which is a further substance known to be under the global regulation of the *gafA* gene in *Pseudomonas fluorescens* (WO 94/01561). The insertion transposon TnCIB116 was used to generate libraries of mutants in MOCG134 and a *gafA*$^-$ derivative of MOCG134 (BL1826). The former was screened for changes in fungal inhibition in vitro; the latter was screened for genes regulated by *gafA* after introduction of *gafA* on a plasmid (see Section C). Selected mutants were characterized by HPLC to assay for production of known compounds such as pyrrolnitrin and 2-hexyl-5-propyl-resorcinol. The HPLC assay enabled a comparison of the novel mutants to the wild-type parental strain. In each case, the HPLC peak corresponding to 2-hexyl-5-propyl-resorcinol was missing in the mutant. The mutant derived from MOCG134 is designated BL1846. The mutant derived from BL1826 is designated BL1911. HPLC for resorcinol follows the same procedure as for pyrrolnitrin (see example 11) except that 100% methanol is applied to the column at 20 min to elute resorcinol.

The resorcinol biosynthetic genes can be cloned from the above-identified mutants in the following manner. Genomic DNA is prepared from the mutants, and clones containing the transposon insertion and adjacent Pseudomonas sequence are obtained by selecting for kanamycin resistant clones (kanamycin resistance is encoded by the transposon). The

cloned *Pseudomonas* sequence is then used as a probe to identify the native sequences from a genomic library of *P. fluorescens* MOCG134. The cloned native genes are likely to represent resorcinol biosynthetic genes.

E. Cloning Soraphen Biosynthetic Genes from *Sorangium*

Soraphen is a polyketide antibiotic produced by the myxobacterium *Sorangium cellulosum*. This compound has broad antifungal activities which make it useful for agricultural applications. In particular, soraphen has activity against a broad range of foliar pathogens.

Example 14: **Isolation of the Soraphen Gene Cluster**

Genomic DNA was isolated from *Sorangium cellulosum* and partially digested with *Sau3A*. Fragments of between 30 and 40 kb were size selected and cloned into the cosmid vector pHC79 (Hohn & Collins, Gene 11: 291-298 (1980)) which had been previously digested with *BamHI* and treated with alkaline phosphatase to prevent self ligation. The cosmid library thus prepared was probed with a 4.6 kb fragment which contains the *graI* region of *Streptomyces violaceoruber* strain Tü22 encoding ORFs 1-4 responsible for the biosynthesis of granaticin in *S. violaceoruber*. Cosmid clones which hybridized to the *graI* probe were identified and DNA was prepared for analysis by restriction digestion and further hybridization. Cosmid p98/1 was identified to contain a 1.8 kb *SalI* fragment which hybridized strongly to the *graI* region; this *SalI* fragment was located within a larger 6.5 kb *PvuI* fragment within the ~40 kb insert of p98/1. Determination of the sequence of part of the 1.8 kb *SalI* insert revealed homology to the acetyltransferase proteins required for the synthesis of erythromycin. Restriction mapping of the cosmid p98/1 was undertaken and generated the map depicted in figure 7. A viable culture of E.coli HB101 comprising cosmid clone 98/1 has been deposited with the Agricultural Research Culture Collection (NRRL) at 1815 N. University Street, Peoria, Illinois 61604 U.S.A. on May 20, 1994, under the accession number NRRL B-21255. The DNA sequence of the soraphen gene cluster is disclosed in SEQ ID NO:6.

Example 15: **Functional Analysis of the Soraphen Gene Cluster**

The regions within p98/1 that encode proteins with a role in the biosynthesis of soraphen were identified through gene disruption experiments. Initially, DNA fragments were derived from cosmid p98/1 by restriction with *PvuI* and cloned into the unique *PvuI* cloning site

(which is within the gene for ampicillin resistance) of the wide host-range plasmid pSUP2021 (Simon *et al.* in: Molecular Genetics of the Bacteria-Plant Interaction (*ed.*: A Puhler), Springer Verlag, Berlin pp 98-106 (1983)). Transformed *E. coli* HB101 was selected for resistance to chloramphenicol, but sensitivity to ampicillin. Selected colonies carrying appropriate inserts were transferred to *Sorangium cellulosum* SJ3 by conjugation using the method described in the published application EP 0 501 921 (to Ciba-Geigy). Plasmids were transferred to *E. coli* ED8767 carrying the helper plasmid pUZ8 (Hedges & Mathew, Plasmid 2: 269-278 (1979)) and the donor cells were incubated with *Sorangium cellulosum* SJ3 cells from a stationary phase culture for conjugative transfer essentially as described in EP 0 501 921 (example 5) and EP *the later app.* (example 2). Selection was on kanamycin, phleomycin and streptomycin. It has been determined that no plasmids tested thus far are capable of autonomous replication in *Sorangium cellulosum*, but rather, integration of the entire plasmid into the chromosome by homologous recombination occurs at a site within the cloned fragment at low frequency. These events can be selected for by the presence of antibiotic resistance markers on the plasmid. Integration of the plasmid at a given site results in the insertion of the plasmid into the chromosome and the concomitant disruption of this region from this event. Therefore, a given phenotype of interest, *i.e.*soraphen production, can be assessed, and disruption of the phenotype will indicate that the DNA region cloned into the plasmid must have a role in the determination of this phenotype.

Recombinant pSUP2021 clones with *PvuI* inserts of approximate size 6.5 kb (pSN105/7), 10 kb (pSN120/10), 3.8 kb (pSN120/43-39) and 4.0 kb (pSN120/46) were selected. The map locations (in kb) of these *PvuI* inserts as shown in Figure 7 are: pSN105/7 - 25.0-31.7, pSN120/10 - 2.5-14.5, pSN120/43-39 - 16.1-20.0, and pSN120/46 - 20.0-24.0. pSN105/7 was shown by digestion with *PvuI* and *SalI* to contain the 1.8 kb fragment referred to above in example 11. Gene disruptions with the 3.8, 4.0, 6.5, and 10 kb PvuI fragments all resulted in the elimination of soraphen production. These results indicate that all of these fragments contain genes or fragments of genes with a role in the production of this compound.

Subsequently gene disruption experiments were performed with two *BglII* fragments derived from cosmid p98/1. These were of size 3.2 kb (map location 32.4-35.6 on Figure 7) and 2.9

kb (map location 35.6-38.5 on Figure 7). These fragments were cloned into the *BamHI* site of plasmid pCIB132 that was derived from pSUP2021 according to Figure 8. The ~5 kb *NotI* fragment of pSUP2021 was excised and inverted, followed by the removal of the ~ 3kb *BamHI* fragment. Neither of these *BglII* fragments was able to disrupt soraphen biosynthesis when reintroduced into *Sorangium* using the method described above. This indicates that the DNA of these fragments has no role in soraphen biosynthesis. Examination of the DNA sequence indicates the presence of a thioesterase domain 5' to, but near the *BglII* site at location 32.4. In addition, there are transcription stop codons immediately after the thioesterase domain which are likely to demarcate the end of the ORF1 coding region. As the 2.9 and 3.2 kb *BglII* fragments are immediately to the right of these sequences it is likely that there are no other genes downstream from ORF1 that are involved in soraphen biosynthesis.

Delineation of the left end of the biosynthetic region required the isolation of two other cosmid clones, pJL1 and pJL3, that overlap p98/1 on the left end, but include more DNA leftwards of p98/1. These were isolated by hybridization with the 1.3 kb BamHI fragment on the extreme left end of p98/1 (map location 0.0-1.3) to the *Sorangium cellulosum* gene library. It should be noted that the *BamHI* site at 0.0 does not exist in the *S. cellulosum* chromosome but was formed as an artifact from the ligation of a *Sau3A* restriction fragment derived from the *Sorangium cellulosum* genome into the *BamHI* cloning site of pHC79. Southern hybridization with the 1.3 kb *BamHI* fragment demonstrated that pJL1 and pJL3 each contain an approximately 12.5 kb *BamHI* fragment that contains sequences common to the 1.3 kb fragment as this fragment is in fact delineated by the *BamHI* site at position 1.3. A viable culture of E.coli HB101 comprising cosmid clone pJL3 has been deposited with the Agricultural Research Culture Collection (NRRL) at 1815 N. University Street, Peoria, Illinois 61604 U.S.A. on May 20, 1994, under the accession number NRRL B-21254. Gene disruption experiments using the 12.5 kb *BamHI* fragment indicated that this fragment contains sequences that are involved in the synthesis of soraphen. Gene disruption using smaller *EcoRV* fragments derived from this region indicated the requirement of this region for soraphen biosynthesis. For example, two *EcoRV* fragments of 3.4 and 1.1 kb located adjacent to the distal *BamHI* site at the left end of the 12.5 kb fragment resulted in a reduction in soraphen biosynthesis when used in gene disruption experiments.

Example 16: **Sequence Analysis of the Soraphen Gene Cluster**

The DNA sequence of the soraphen gene cluster was determined from the *PvuI* site at position 2.5 to the *BglII* site at position 32.4 (see Figure 7) using the Taq DyeDeoxy Terminator Cycle Sequencing Kit supplied by Applied Biosystems, Inc., Foster City, CA. following the protocol supplied by the manufacturer. Sequencing reactions were run on a Applied Biosystems 373A Automated DNA Sequencer and the raw DNA sequence was assembled and edited using the "INHERIT" software package also from Applied Biosystems, Inc.. The pattern recognition program "FRAMES" was used to search for open reading frames (ORFs) in all six translation frames of the DNA sequence. In total approximately 30 kb of contiguous DNA was assembled and this corresponds to the region determined to be critical to soraphen biosynthesis in the disruption experiments described in example 12. This sequence encodes two ORFs which have the structure described below.

ORF1:

ORF1 is approximately 25.5 kb in size and encodes five biosynthetic modules with homology to the modules found in the erythromycin biosynthetic genes of *Saccharopolyspora erythraea* (Donadio *et al.* Science <u>252</u>: 675-679 (1991)). Each module contains a β-ketoacylsynthase (KS), an acyltransferase (AT), a ketoreductase (KR) and an acyl carrier protein (ACP) domain as well as β-ketone processing domains which may include a dehydratase (DH) and/or enoyl reductase (ER) domain. In the biosynthesis of the polyketide structure each module directs the incorporation of a new two carbon extender unit and the correct processing of the β-ketone carbon.

ORF2:

In addition to ORF1, DNA sequence data from the p98/1 fragment spanning the *PvuI* site at 2.5 kb and the *SmaI* site at 6.2 kb, indicated the presence of a further ORF (ORF2) immediately adjacent to ORF1. The DNA sequence demonstrates the presence of a typical biosynthetic module that appears to be encoded on an ORF whose 5' end is not yet sequenced and is some distance to the left. By comparison to other polyketide biosynthetic gene units and the number of carbon atoms in the soraphen ring structure it is likely that there should be a total of eight modules in order to direct the synthesis of 17 carbon molecule soraphen. Since there are five modules in ORF1 described above, it was predicted that ORF2 contains a further three and that these would extend beyond the left

end of cosmid p98/1 (position 0 in Figure 7). This is entirely consistent with the gene description of example 12. The cosmid clones pJL1 and pJL3 extending beyond the left end of p98/1 presumable carry the sequence encoding the remaining modules required for soraphen biosynthesis.

Example 17: **Soraphen: Requirement for Methylation**

Synthesis of polyketides typically requires, as a first step, the condensation of a starter unit (commonly acetate) and an extender unit (malonate) with the loss of one carbon atom in the form of CO_2 to yield a three-carbon chain. All subsequent additions result in the addition of two carbon units to the polyketide ring (Donadio *et al.* Science 252: 675-679 (1991)). Since soraphen has a 17-carbons ring, it is likely that there are 8 biosynthetic modules required for its synthesis. Five modules are encoded in ORF1 and a sixth is present at the 3' end of ORF2. As explained above, it is likely that the remaining two modules are also encoded by ORF2 in the regions that are in the 15 kb *BamHI* fragment from pJL1 and pJL3 for which the sequence has not yet been determined.

The polyketide modular biosynthetic apparatus present in *Sorangium cellulosum* is required for the production of the compound, soraphen C, which has no antipathogenic activity. The structure of this compound is the same as that of the antipathogenic soraphen A with the exception that the O-methyl groups of soraphen A at positions 6, 7, and 14 of the ring are hydroxyl groups. These are methylated by a specific methyltransferase to form the active compound soraphen A. A similar situation exists in the biosynthesis of erythromycin in *Saccharopolyspora erythraea*. The final step in the biosynthesis of this molecule is the methylation of three hydroxl groups by a methyltransferase (Haydock *et al.*, Mol. Gen. Genet. 230: 120-128 (1991)). It is highly likely, therefore, that a similar methyltransferase (or possibly more than one) operates in the biosynthesis of soraphen A (soraphen C is unmethylated and soraphen B is partially methylated). In all polyketide biosynthesis systems examined thus far, all of the biosynthetic genes and associated methylases are clustered together (Summers *et al.* J Bacteriol 174: 1810-1820 (1992)). It is also probable, therefore, that a similar situation exists in the soraphen operon and that the gene encoding the methyltransferase/s required for the conversion of soraphen B and C to soraphen A is located near the ORF1 and ORF2 that encode the polyketide synthase. The results of the gene disruption experiments described above indicate that this gene is not located

immediately downstream from the 3' end of ORF1 and that it is likely located upstream of ORF2 in the DNA contained in pJL1 and pJL3. Thus, using standard techniques in the art, the methyltransferase gene can be cloned and sequenced.

Soraphen Determination

Sorangium cellulosum cells were cultured in a liquid growth medium containing an exchange resin, XAD-5 (Rohm and Haas) (5% w/v). The soraphen A produced by the cells bound to the resin which was collected by filtration through a polyester filter (Sartorius B 420-47-N) and the soraphen was released from the resin by extraction with 50 ml isopropanol for 1 hr at 30 C. The isopropanol containing soraphen A was collected and concentrated by drying to a volume of approximately 1 ml. Aliquots of this sample were analyzed by HPLC at 210 nm to detect and quantify the soraphen A. This assay procedure is specific for soraphen A (fully methylated); partially and non-methylated soraphen forms have a different R_T and are not measured by this procedure. This procedure was used to assay soraphen A production after gene disruption.

F. Cloning and Characterization of Phenazine Biosynthetic Genes from *Pseudomonas aureofaciens*

The phenazine antibiotics are produced by a variety of *Pseudomonas* and *Streptomyces* species as secondary metabolites branching off the shikimic acid pathway. It has been postulated that two chorismic acid molecules are condensed along with two nitrogens derived from glutamine to form the three-ringed phenazine pathway precursor phenazine-1,6-dicarboxylate. However, there is also genetic evidence that anthranilate is an intermediate between chorismate and phenazine-1,6-dicarboxylate (Essar *et al.*, J. Bacteriol. 172: 853-866 (1990)). In *Pseudomonas aureofaciens* 30-84, production of three phenazine antibiotics, phenazine-1-carboxylic acid, 2-hydroxyphenazine-1-carboxylic acid, and 2-hydroxyphenazine, is the major mode of action by which the strain protects wheat from the fungal phytopathogen *Gaeumannomyces graminis* var. *tritici* (Pierson & Thomashow, MPMI 5: 330-339 (1992)). Likewise, in *Pseudomonas fluorescens* 2-79, phenazine production is a major factor in the control of *G. graminis* var. *tritici* (Thomashow & Weller, J. Bacteriol. 170: 3499-3508 (1988)).

Example 18: **Isolation of the Phenazine Biosynthetic Genes**

Pierson & Thomashow (*supra*) have previously described the cloning of a cosmid which confers a phenazine biosynthesis phenotype on transposon insertion mutants of *Pseudomonas aureofaciens* strain 30-84 which were disrupted in their ability to synthesize phenazine antibiotics. A mutant library of strain 30-84 was made by conjugation with *E. coli* S17-1(pSUP1021) and mutants unable to produce phenazine antibiotics were selected. Selected mutants were unable to produce phenazine carboxylic acid, 2-hydroxyphenaxine or 2-hydroxy-phenazine carboxylic acid. These mutants were transformed by a cosmid genomic library of strain 30-84 leading to the isolation of cosmid pLSP259 which had the ability to complement phenazine mutants by the synthesis of phenazine carboxylic acid, 2-hydroxyphenazine and 2-hydroxy-phenazinecarboxylic acid. pLSP259 was further characterized by transposon mutagenesis using the λ::Tn5 phage described by de Bruijn & Lupski (Gene 27: 131-149 (1984)). Thus a segment of approximately 2.8 kb of DNA was identified as being responsible for the phenazine complementing phenotype; this 2.8 kb segment is located within a larger 9.2 kb *EcoRI* fragment of pLSP259. Transfer of the 9.2 kb *EcoRI* fragment and various deletion derivatives thereof to *E. coli* under the control of the *lacZ* promoter was undertaken to assay for the production in *E. coli* of phenazine. The shortest deletion derivative which was found to confer biosynthesis of all three phenazine compounds to *E. coli* contained an insert of approximately 6 kb and was designated pLSP18-6H3del3. This plasmid contained the 2.8 kb segment previously identified as being critical to phenazine biosynthesis in the host 30-84 strain and was provided by Dr LS Pierson (Department of Plant Pathology, U Arizona, Tucson, AZ) for sequence characterization. Other deletion derivatives were able to confer production of phenazine-carboxylic acid on *E. coli*, without the accompanying production of 2-hydroxyphenazine and 2-hydroxyphenazinecarboxylic acid suggesting that at least two genes might be involved in the synthesis of phenazine and its hydroxy derivatives.

The DNA sequence comprising the genes for the biosynthesis of phenazine is disclosed in SEQ ID NO:17. Plasmid pCIB3350 contains the PstI-HindIII fragment of the phenazine gene cluster and has been deposited with the Agricultural Research Culture Collection (NRRL) at 1815 N. University Street, Peoria, Illinois 61604 U.S.A. on May 20, 1994, under the accession number NRRL B-21257. Plasmid pCIB3351 contains the EcoRI-PstI fragment of the phenazine gene cluster and has been deposited with the Agricultural Research Culture

Collection (NRRL) at 1815 N. University Street, Peoria, Illinois 61604 U.S.A. on May 20, 1994, under the accession number NRRL B-21258. pCIB3350 along with pCIB3351 comprises the entire phenazine gene of SEQ ID NO:17. Determination of the DNA sequence of the insert of pLSP18-6H3del3 revealed the presence of four ORFs within and adjacent to the critical 2.8 kb segment. ORF1 (SEQ ID NO:18) was designated *phz1*, ORF2 (SEQ ID NO:19) was designated *phz2*, and ORF3 (SEQ ID NO:20) was designated *phz3*, and ORF4 (SEQ ID NO:22) was designated *phz4*. The DNA sequence of *phz4* is shown in SEQ ID NO:21. *phz1* is approximately 1.35 kb in size and has homology at the 5' end to the entB gene of *E. coli*, which encodes isochorismatase. *phz2* is approximately 1.15 kb in size and has some homology at the 3' end to the trpG gene which encodes the beta subunit of anthranilate synthase. *phz3* is approximately 0.85 kb in size. *phz4* is approximately 0.65 kb in size and is homologous to the pdxH gene of *E. coli* which encodes pyridoxamine 5'-phosphate oxidase.

Phenazine Determination

Thomashow *et al.* (Appl Environ Microbiol 56: 908-912 (1990)) describe a method for the isolation of phenazine. This involves acidifying cultures to pH 2.0 with HCl and extraction with benzene. Benzene fractions are dehydrated with Na_2SO_4 and evaporated to dryness. The residue is redissolved in aqueous 5% $NaHCO_3$, reextracted with an equal volume of benzene, acidified, partitioned into benzene and redried. Phenazine concentrations are determined after fractionation by reverse-phase HPLC as described by Thomashow *et al.* (*supra*).

G. Cloning Peptide Antipathogenic Genes

This group of substances is diverse and is classifiable into two groups: (1) those which are synthesized by enzyme systems without the participation of the ribosomal apparatus, and (2) those which require the ribosomally-mediated translation of an mRNA to provide the precursor of the antibiotic.

Non-Ribosomal Peptide Antibiotics.

Non-Ribosomal Peptide Antibiotics are assembled by large, multifunctional enzymes which activate, modify, polymerize and in some cases cyclize the subunit amino acids, forming

polypeptide chains. Other acids, such as aminoadipic acid, diaminobutyric acid, diaminopropionic acid, dihydroxyamino acid, isoserine, dihydroxybenzoic acid, hydroxyisovaleric acid, (4R)-4-[(E)-2-butenyl]-4,N-dimethyl-L-threonine, and ornithine are also incorporated (Katz & Demain, Bacteriological Review 41: 449-474 (1977); Kleinkauf & von Dohren, Annual Review of Microbiology 41: 259-289 (1987)). The products are not encoded by any mRNA, and ribosomes do not directly participate in their synthesis. Peptide antibiotics synthesized non-ribosomally can in turn be grouped according to their general structures into linear, cyclic, lactone, branched cyclopeptide, and depsipeptide categories (Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990)). These different groups of antibiotics are produced by the action of modifying and cyclizing enzymes; the basic scheme of polymerization is common to them all. Non-ribosomally synthesized peptide antibiotics are produced by both bacteria and fungi, and include edeine, linear gramicidin, tyrocidine and gramicidin S from *Bacillus brevis*, mycobacillin from *Bacillus subtilis*, polymyxin from *Bacillus polymiyxa*, etamycin from *Streptomyces griseus*, echinomycin from *Streptomyces echinatus*, actinomycin from *Streptomyces clavuligerus*, enterochelin from *Escherichia coli*, gamma-(alpha-L-aminoadipyl)-L-cysteinyl-D-valine (ACV) from *Aspergillus nidulans*, alamethicine from *Trichoderma viride*, destruxin from *Metarhizium anisolpliae*, enniatin from *Fusarium oxysporum*, and beauvericin from *Beauveria bassiana*. Extensive functional and structural similarity exists between the prokaryotic and eukaryotic systems, suggesting a common origin for both. The activities of peptide antibiotics are similarly broad, toxic effects of different peptide antibiotics in animals, plants, bacteria, and fungi are known (Hansen, Annual Review of Microbiology 47: 535-564 (1993); Katz & Demain, Bacteriological Reviews 41: 449-474 (1977); Kleinkauf & von Dohren, Annual Review of Microbiology 41: 259-289 (1987); Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990); Kolter & Moreno, Annual Review of Microbiology 46: 141-163 (1992)).

Amino acids are activated by the hydrolysis of ATP to form an adenylated amino or hydroxy acid, analogous to the charging reactions carried out by aminoacyl-tRNA synthetases, and then covalent thioester intermediates are formed between the amino acids and the enzyme(s), either at specific cysteine residues or to a thiol donated by pantetheine. The amino acid-dependent hydrolysis of ATP is often used as an assay for peptide antibiotic enzyme complexes (Ishihara, *et al.*, Journal of Bacteriology 171: 1705-1711 (1989)). Once

bound to the enzyme, activated amino acids may be modified before they are incorporated into the polypeptide. The most common modifications are epimerization of L-amino (hydroxy) acids to the D- form, N-acylations, cyclizations and N-methylations. Polymerization occurs through the participation of a pantetheine cofactor, which allows the activated subunits to be sequentially added to the polypeptide chain. The mechanism by which the peptide is released from the enzyme complex is important in the determination of the structural class in which the product belongs. Hydrolysis or aminolysis by a free amine of the thiolester will yield a linear (unmodified or terminally aminated) peptide such as edeine; aminolysis of the thiolester by amine groups on the peptide itself will give either cyclic (attack by terminal amine), such as gramicidin S, or branched (attack by side chain amine), such as bacitracin, peptides; lactonization with a terminal or side chain hydroxy will give a lactone, such as destruxin, branched lactone, or cyclodepsipeptide, such as beauvericin.

The enzymes which carry out these reactions are large multifunctional proteins, having molecular weights in accord with the variety of functions they perform. For example, gramicidin synthetases 1 and 2 are 120 and 280 kDa, respectively; ACV synthetase is 230 kDa; enniatin synthetase is 250 kDa; bacitracin synthetases 1, 2, 3 are 335, 240, and 380 kDa, respectively (Katz & Demain, Bacteriological Reviews 41: 449-474 (1977); Kleinkauf & von Dohren, Annual Review of Microbiology 41: 259-289 (1987); Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990). The size and complexity of these proteins means that relatively few genes must be cloned in order for the capability for the complete nonribosomal synthesis of peptide antibiotics to be transferred. Further, the functional and structural homology between bacterial and eukaryotic synthetic systems indicates that such genes from any source of a peptide antibiotic can be cloned using the available sequence information, current functional information, and conventional microbiological techniques. The production of a fungicidal, insecticidal, or batericidal peptide antibiotic in a plant is expected to produce an advantage with respect to the resistance to agricultural pests.

Example 19: **Cloning of Gramicidin S Biosynthesis Genes**

Gramicidin S is a cyclic antibiotic peptide and has been shown to inhibit the germination of fungal spores (Murray, *et al.*, Letters in Applied Microbiology 3: 5-7 (1986)), and may

therefore be useful in the protection of plants against fungal diseases. The gramicidin S biosynthesis operon (*grs*) from *Bacillus brevis* ATCC 9999 has been cloned and sequenced, including the entire coding sequences for gramicidin synthetase 1 (GS1, *grsA*), another gene in the operon of unknown function (*grsT*), and GS2 (*grsB*) (Kratzschmar, *et al.*, Journal of Bacteriology 171: 5422-5429 (1989); Krause, *et al.*, Journal of Bacteriology 162: 1120-1125 (1985)). By methods well known in the art, pairs of PCR primers are designed from the published DNA sequence which are suitable for amplifying segments of approximately 500 base pairs from the grs operon using isolated *Bacillus brevis* ATCC 9999 DNA as a template. The fragments to be amplified are (1) at the 3' end of the coding region of *grsB*, spanning the termination codon, (2) at the 5' end of the *grsB* coding sequence, including the initiation codon, (3) at the 3' end of the coding sequence of *grsA*, including the termination codon, (4) at the 5' end of the coding sequence of *grsA*, including the initiation codon, (5) at the 3' end of the coding sequence of *grsT*, including the termination codon, and (6) at the 5' end of the coding sequence of *grsT*, including the initiation codon. The amplified fragments are radioactively or nonradioactively labeled by methods known in the art and used to screen a genomic library of *Bacillus brevis* ATCC 9999 DNA constructed in a vector such as λEMBL3. The 6 amplified fragments are used in pairs to isolate cloned fragments of genomic DNA which contain intact coding sequences for the three biosynthetic genes. Clones which hybridize to probes 1 and 2 will contain an intact *grsB* sequence, those which hybridize to probes 3 and 4 will contain an intact *grsA* gene, those which hybridize to probes 5 and 6 will contain an intact *grsT* gene. The cloned *grsA* is introduced into *E. coli* and extracts prepared by lysing transformed bacteria through methods known in the art are tested for activity by the determination of phenylalanine-dependent ATP-PP_i exchange (Krause, *et al.*, Journal of Bacteriology 162: 1120-1125 (1985)) after removal of proteins smaller than 120 kDa by gel filtration chromatography. *GrsB* is tested similarly by assaying gel-filtered extracts from transformed bacteria for proline, valine, ornithine and leucine-dependent ATP-PP_i exchange.

Example 20: **Cloning of Penicillin Biosynthesis Genes**

A 38 kb fragment of genomic DNA from *Penicillium chrysogenum* transfers the ability to synthesize penicillin to fungi, *Aspergillus niger*, and *Neurospora crassa*, which do not normally produce it (Smith, *et al.*, Bio/Technology 8: 39-41 (1990)). The genes which are responsible for biosynthesis, delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase,

isopenicillin N synthetase, and isopenicillin N acyltranferase have been individually cloned from *P. chrysogenum* and *Aspergillus nidulans*, and their sequences determined (Ramon, *et al.*, Gene 57: 171-181 (1987); Smith, *et al.*, EMBO Journal 9: 2743-2750 (1990); Tobin, *et al.*, Journal of Bacteriology 172: 5908-5914 (1990)). The cloning of these genes is accomplished by following the PCR-based approach described above to obtain probes of approximately 500 base pairs from genomic DNA from either *Penicillium chrysogenum* (for example, strain AS-P-78, from Antibioticos, S.A., Leon, Spain), or from *Aspergillus nidulans* for example, strain G69. Their integrity and function may be checked by transforming the non-producing fungi listed above and assaying for antibiotic production and individual enzyme activities as described (Smith, *et al.*, Bio/Technology 8: 39-41 (1990)).

Example 21: **Cloning of Bacitracin A Biosynthesis Genes**

Bacitracin A is a branched cyclopeptide antibiotic which has potential for the enhancement of disease resistance to bacterial plant pathogens. It is produced by *Bacillus licheniformis* ATCC 10716, and three multifunctional enzymes, bacitracin synthetases (BA) 1, 2, and 3, are required for its synthesis. The molecular weights of BA1, BA2, and BA3 are 335 kDa, 240 kDa, and 380 kDa, respectively. A 32 kb fragment of *Bacillus licheniformis* DNA which encodes the BA2 protein and part of the BA3 protein shows that at least these two genes are linked (Ishihara, *et al.*, Journal of Bacteriology 171: 1705-1711 (1989)). Evidence from gramicidin S, penicillin, and surfactin biosynthetic operons suggest that the first protein in the pathway, BA1, will be encoded by a gene which is relatively close to BA2 and BA3. BA3 is purified by published methods, and it is used to raise an antibody in rabbits (Ishihara, *et al. supra*). A genomic library of *Bacillus licheniformis* DNA is transformed into *E. coli* and clones which express antigenic determinants related to BA3 are detected by methods known in the art. Because BA1, BA2, and BA3 are antigenically related, the detection method will provide clones encoding each of the three enzymes. The identity of each clone is confirmed by testing extracts of transformed *E. coli* for the appropriate amino acid-dependent ATP-PP_i exchange. Clones encoding BA1 will exhibit leucine-, glutamic acid-, and isoleucine-dependent ATP-PP_i exchange, those encoding BA2 will exhibit lysine- and ornithine-dependent exchange, and those encoding BA3 will exhibit isoleucine, phenylalanine-, histidine-, aspartic acid-, and asparagine-dependent exchange. If one or two genes are obtained by this method, the others are isolated by techniques known in the art as "walking"

or "chromosome walking" techniques (Sambrook et al, in: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labroatory Press, 1989).

Example 22: **Cloning of Beauvericin and Destruxin Biosynthesis Genes**

Beauvericin is an insecticidal hexadepsipeptide produced by the fungus *Beauveria bassiana* (Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990)) which will provide protection to plants from insect pests. It is an analog of enniatin, a phytotoxic hexadepsipeptide produced by some phytopathogenic species of *Fusarium* (Burmeister & Plattner, Phytopathology 77: 1483-1487 (1987)). Destruxin is an insecticidal lactone peptide produced by the fungus *Metarhizium anisopliae* (James, *et al.*, Journal of Insect Physiology 39: 797-804 (1993)). Monoclonal antibodies directed to the region of the enniatin synthetase complex responsible for N-methylation of activated amino acids cross react with the synthetases for beauvericin and destruxin, demonstrating their structural relatedness (Kleinkauf & von Dohren, European Journal of Biochemistry 192: 1-15 (1990)). The gene for enniatin synthetase gene (*esyn1*) from *Fusarium scirpi* has been cloned and sequenced (Haese, *et al.*, Molecular Microbiology 7: 905-914 (1993)), and the sequence information is used to carry out a cloning strategy for the beauvericin synthetase and destruxin synthetase genes as described above. Probes for the beauvericin synthetase (BE) gene and the destruxin synthetase (DXS) gene are produced by amplifying specific regions of *Beauveria bassiana* genomic DNA or *Metarhizium anisopliae* genomic DNA using oligomers whose sequences are taken from the enniatin synthetase sequence as PCR primers. Two pairs of PCR primers are chosen, with one pair capable of causing the amplification of the segment of the BE gene spanning the initiation codon, and the other pair capable of causing the amplification of the segment of the BE gene which spans the termination codon. Each pair will cause the production of a DNA fragment which is approximately 500 base pairs in size. Library of genomic DNA from *Beauveria bassiana* and *Metarhizium anisopliae* are probed with the labeled fragments, and clones which hybridize to both of them are chosen. Complete coding sequences of beauvericin synthetase will cause the appearance of phenylalanine-dependent ATP-PP_i exchange in an appropriate host, and that of destruxin will cause the appearance of valine-, isoleucine-, and alanine-dependent ATP-PP_i exchange. Extracts from these transformed organisms will also carry out the cell-free biosynthesis of beauvericin and destruxin, respectively.

Example 23: **Cloning genes for the Biosynthesis of an Unknown Peptide Antibiotic**

The genes for any peptide antibiotic are cloned by the use of conserved regions within the coding sequence. The functions common to all peptide antibiotic synthetases, that is, amino acid activation, ATP-, and pantotheine binding, are reflected in a repeated domain structure in which each domain spans approximately 600 amino acids. Within the domains, highly conserved sequences are known, and it is expected that related sequences will exist in any peptide antibiotic synthetase, regardless of its source. The published DNA sequences of peptide synthetase genes, including gramicidin synthetases 1 and 2 (Hori, *et al.*, Journal of Biochemistry 106: 639-645 (1989); Krause, *et al.*, Journal of Bacteriology 162: 1120-1125 (1985); Turgay, *et al.*, Molecular Microbiology 6: 529-546 (1992)), tyrocidine sythethase 1 and 2 (Weckermann, *et al.*, Nucleic Acids Research 16: 11841 (1988)), ACV synthetase (MacCabe, *et al.*, Journal of Biological Chemistry 266: 12646-12654 (1991)), enniatin synthetase (Haese, *et al.*, Molecular Microbiology 7: 905-914 (1993)), and surfactin synthetase (Fuma, *et al.*, Nucleic Acids Research 21: 93-97 (1993); Grandi, *et al.*, Eleventh International Spores Conference (1992)) are compared and the individual repeated domains are identified. The domains from all the synthetases are compared as a group, and the most highly conserved sequences are identified. From these conserved sequences, DNA oligomers are designed which are suitable for hybridizing to all of the observed variants of the sequence, and another DNA sequence which lies, for example, from 0.1 to 2 kilobases away from the first DNA sequence, is used to design another DNA oligomer. Such pairs of DNA oligomers are used to amplify by PCR the intervening segment of the unknown gene by combining them with genomic DNA prepared from the organism which produces the antibiotic, and following a PCR amplification procedure. The fragment of DNA which is produced is sequenced to confirm its identity, and used as a probe to identify clones containing larger segments of the peptide synthetase gene in a genomic library. A variation of this approach, in which the oligomers designed to hybridize to the conserved sequences in the genes were used as hybridization probes themselves, rather than as primers of PCR reactions, resulted in the identification of part of the surfactin synthetase gene from *Bacillus subtilis* ATCC 21332 (Borchert, *et al.*, FEMS Microbiological Letters 92: 175-180 (1992)). The cloned genomic DNA which hybridizes to the PCR-generated probe is sequenced, and the complete coding sequence is obtained by "walking" procedures. Such "walking" procedures will also yield other genes required for the peptide antibiotic synthesis, because they are known to be clustered.

Another method of obtaining the genes which code for the synthetase(s) of a novel peptide antibiotic is by the detection of antigenic determinants expressed in a heterologous host after transformation with an appropriate genomic library made from DNA from the antibiotic-producing organism. It is expected that the common structural features of the synthetases will be evidenced by cross-reactions with antibodies raised against different synthetase proteins. Such antibodies are raised against peptide synthetases purified from known antibiotic-producing organisms by known methods (Ishihara, *et al.*, Journal of Bacteriology 171: 1705-1711 (1989)). Transformed organisms bearing fragments of genomic DNA from the producer of the unknown peptide antibiotic are tested for the presence of antigenic determinants which are recognized by the anti-peptide synthetase antisera by methods known in the art. The cloned genomic DNA carried by cells which are identified by the antisera are recovered and sequenced. "Walking" techniques, as described earlier, are used to obtain both the entire coding sequence and other biosynthetic genes.

Another method of obtaining the genes which code for the synthetase of an unknown peptide antibiotic is by the purification of a protein which has the characteristics of the appropriate peptide synthetase, and determining all or part of its amino acid sequence. The amino acids present in the antibiotic are determined by first purifying it from a chloroform extract of a culture of the antibiotic-producing organism, for example by reverse phase chromatography on a C_{18} column in an ethanol-water mixture. The composition of the purified compound is determined by mass spectrometry, NMR, and analysis of the products of acid hydrolysis. The amino or hydroxy acids present in the peptide antibiotic will produce ATP-PP_i exchange when added to a peptide-synthetase-containing extract from the antibiotic-producing organism. This reaction is used as an assay to detect the presence of the peptide synthetase during the course of a protein purification scheme, such as are known in the art. A substantially pure preparation of the peptide synthetase is used to determine its amino acid sequence, either by the direct sequencing of the intact protein to obtain the N-terminal amino acid sequence, or by the production, purification, and sequencing of peptides derived from the intact peptide synthetase by the action of specific proteolytic enzymes, as are known in the art. A DNA sequence is inferred from the amino acid sequence of the synthetase, and DNA oligomers are designed which are capable of hybridizing to such a coding sequence. The oligomers are used to probe a genomic library

made from the DNA of the antibiotic-producing organism. Selected clones are sequenced to identify them, and complete coding sequences and associated genes required for peptide biosynthesis are obtained by using "walking" techniques. Extracts from organisms which have been transformed with the entire complement of peptide biosynthetic genes, for example bacteria or fungi, will produce the peptide antibiotic when provided with the required amino or hydroxy acids, ATP, and pantetheine.

Further methods appropriate for the cloning of genes required for the synthesis of non-ribosomal peptide antibiotics are described in Section B of the examples.

Ribosomally-Synthesized Peptide Antibiotics.

Ribosomally-Synthesized Peptide Antibiotics are characterized by the existence of a structural gene for the antibiotic itself, which encodes a precursor that is modified by specific enzymes to create the mature molecule. The use of the general protein synthesis apparatus for peptide antibiotic synthesis opens up the possibility for much longer polymers to be made, although these peptide antibiotics are not necessarily very large. In addition to a structural gene, further genes are required for extracellular secretion and immunity, and these genes are believed to be located close to the structural gene, in most cases probably on the same operon. Two major groups of peptide antibiotics made on ribosomes exist: those which contain the unusual amino acid lanthionine, and those which do not. Lanthionine-containing antibiotics (lantibiotics) are produced by gram-positive bacteria, including species of *Lactococcus*, *Staphylococcus*, *Streptococcus*, *Bacillus*, and *Streptomyces*. Linear lantibiotics (for example, nisin, subtilin, epidermin, and gallidermin), and circular lantibiotics (for example, duramycin and cinnamycin), are known (Hansen, Annual Review of Microbiology 47: 535-564 (1993); Kolter & Moreno, Annual Review of Microbiology 46: 141-163 (1992)). Lantibiotics often contain other characteristic modified residues such as dehydroalanine (DHA) and dehydrobutyrine (DHB), which are derived from the dehydration of serine and threonine, respectively. The reaction of a thiol from cysteine with DHA yields lanthionine, and with DHB yields β-methyllanthionine. Peptide antibiotics which do not contain lanthionine may contain other modifications, or they may consist only of the ordinary amino acids used in protein synthesis. Non-lanthionine-containing peptide antibiotics are produced by both gram-positive and gram-negative bacteria, including *Lactobacillus*, *Lactococcus*, *Pediococcus*, *Enterococcus*, and

Escherichia. Antibiotics in this category include lactacins, lactocins, sakacin A, pediocins, diplococcin, lactococcins, and microcins (Hansen, *supra*; Kolter & Moreno, *supra*). In general, peptide antibiotics whose synthesis is begun on ribosomes are subject to several types of post-translational processing, including proteolytic cleavage and modification of amino acid side chains, and require the presence of a specific transport and/or immunity mechanism. The necessity for protection from the effects of these antibiotics appears to contrast strongly with the lack of such systems for nonribosomal peptide antibiotics. This may be rationalized by considering that the antibiotic activity of many ribosomally-synthesized peptide antibiotics is directed at a narrow range of bacteria which are fairly closely related to the producing organism. In this situation, a particular method of distinguishing the producer from the competitor is required, or else the advantage is lost. As antibiotics, this property has limited the usefulness of this class of molecules for situations in which a broad range of activity if desirable, but enhances their attractiveness in cases when a very limited range of activities is advantageous. In eukaryotic systems, which are not known to be sensitive to any of this type of peptide antibiotic, it is not clear if production of a ribosomally-synthesized peptide antibiotic necessitates one of these transport systems, or if transport out of the cell is merely a matter of placing the antibiotic in a better location to encounter potential pathogens. This question can be addressed experimentally, as shown in the examples which follow.

Example 24: **Cloning Genes for the Biosynthesis of a Lantibiotic**

Examination of genes linked to the structural genes for the lantibiotics nisin, subtilin, and epidermin show several open reading frames which share sequence homology, and the predicted amino acid sequences suggest functions which are necessary for the maturation and transport of the antibiotic. The *spa* genes of *Bacillus subtilis* ATCC 6633, including *spaS*, the structural gene encoding the precursor to subtilin, have been sequenced (Chung & Hansen, Journal of Bacteriology 174: 6699-6702 (1992); Chung, *et al.*, Journal of Bacteriology 174: 1417-1422 (1992); Klein, *et al.*, Applied and Environmental Microbiology 58: 132-142 (1992)). Open reading frames were found only upstream of *spaS*, at least within a distance of 1-2 kilobases. Several of the open reading frames appear to part of the same transcriptional unit, *spaE*, *spaD*, *spaB*, and *spaC*, with a putative promoter upstream of *spaE*. Both *spaB*, which encodes a protein of 599 amino acids, and *spaD*, which encodes a protein of 177 amino acids, share homology to genes required for the transport

of hemolysin, coding for the HylB and HlyD proteins, respectively. *SpaE*, which encodes a protein of 851 amino acids, is homologous to *nisB*, a gene linked to the structural gene for nisin, for which no function is known. *SpaC* codes for a protein of 442 amino acids of unknown function, but disruption of it eliminates production of subtilin. These genes are contained on a segment of genomic DNA which is approximately 7 kilobases in size (Chung & Hansen, Journal of Bacteriology 174: 6699-6702 (1992); Chung, *et al.*, Journal of Bacteriology 174: 1417-1422 (1992); Klein, *et al.*, Applied and Environmental Microbiology 58: 132-142 (1992)). It has not been clearly demonstrated if these genes are completely sufficient to confer the ability to produce subtilin. A 13.5 kilobasepair (kb) fragment from plasmid Tü32 of *Staphylococcus epidermis* Tü3298 containing the structural gene for epidermin (*epiA*), also contains five open reading frames denoted *epiA*, *epiB*, *epiC*, *epiD*, *epiQ*, and *epiP*. The genes *epiBC* are homologous to the genes *spaBC*, while *epiQ* appears to be involved in the regulation of the expression of the operon, and *epiP* may encode a protease which acts during the maturation of pre-epidermin to epidermin. *EpiD* encodes a protein of 181 amino acids which binds the coenzyme flavin mononucleotide, and is suggested to perform post-translational modification of pre-epidermin (Kupke, *et al.*, Journal of Bacteriology 174: (1992); Peschel, *et al.*, Molecular Microbiology 9: 31-39 (1993); Schnell, *et al.*, European Journal of Biochemistry 204: 57-68 (1992)). It is expected that many, if not all, of the genes required for the biosynthesis of a lantibiotic will be clustered, and physically close together on either genomic DNA or on a plasmid, and an approach which allows one of the necessary genes to be located will be useful in finding and cloning the others. The structural gene for a lantibiotic is cloned by designing oligonucleotide probes based on the amino acid sequence determined from a substantially purified preparation of the lantibiotic itself, as has been done with the lantibiotics lacticin 481 from *Lactococcus lactis* subsp. *lactis* CNRZ 481 (Piard, *et al.*, Journal of Biological Chemistry 268: 16361-16368 (1993)), streptococcin A-FF22 from *Streptococcus pyogenes* FF22 (Hynes, *et al.*, Applied and Environmental Microbiology 59: 1969-1971 (1993)), and salivaricin A from *Streptococcus salivarius* 203P (Ross, *et al.*, Applied and Environmental Microbiology 59: 2014-2021 (1993)). Fragments of bacterial DNA approximately 10-20 kilobases in size containing the structural gene are cloned and sequenced to determine regions of homology to the characterized genes in the *spa*, *epi*, and *nis* operons. Open reading frames which have homology to any of these genes or which lie in the same transcriptional unit as open reading frames having homology to any of these genes are

cloned individually using techniques known in the art. A fragment of DNA containing all of the associated reading frames and no others is transformed into a non-producing strain of bacteria, such as *Esherichia coli*, and the production of the lantibiotic analyzed, in order to demonstrate that all the required genes are present.

Example 25: **Cloning Genes for the Biosynthesis of a Non-Lanthionine Containing, Ribosomally Synthesized Peptide Antibiotic**

The lack of the extensive modifications present in lantibiotics is expected to reduce the number of genes required to account for the complete synthesis of peptide antibiotics exemplified by lactacin F, sakacin A, lactococcin A, and helveticin J. Clustered genes involved in the biosynthesis of antibiotics were found in *Lactobacillus johnsonii* VPI11088, for lactacin F (Fremaux, *et al.*, Applied and Environmental Microbiology 59: 3906-3915 (1993)), in *Lactobacillus sake* Lb706 for sakacin A (Axelsson, *et al.*, Applied and Environmental Microbiology 59: 2868-2875 (1993)), in *Lactococcus lactis* for lactococcin A (Stoddard, *et al.*, Applied and Environmental Microbiology 58: 1952-1961 (1992)), and in *Pediococcus acidilactici* for pediocin PA-1 (Marugg, *et al.*, Applied and Environmental Microbiology, 58: 2360-2367 (1992)). The genes required for the biosynthesis of a novel non-lanthionine-containing peptide antibiotic are cloned by first determining the amino acid sequence of a substantially purified preparation of the antibiotic, designing DNA oligomers based on the amino acid sequence, and probing a DNA library constructed from either genomic or plasmid DNA from the producing bacterium. Fragments of DNA of 5-10 kilobases which contain the structural gene for the antibiotic are cloned and sequenced. Open reading frames which have homology to *sakB* from *Lactobacillus sake*, or to lafX, ORFY, or ORFZ from *Lactobacillus johnsonii*, or which are part of the same transcriptional unit as the antibiotic structural gene or genes having homology to those genes previously mentioned are individually cloned by methods known in the art. A fragment of DNA containing all of the associated reading frames and no others is transformed into a non-producing strain of bacteria, such as Esherichia coli,, and the production of the antibiotic analyzed, in order to demonstrate that all the required genes are present.

H. Expression of Antibiotic Biosynthetic Genes in Microbial Hosts

Example 26: Overexpression of APS Biosynthetic Genes for Overproduction of APS using Fermentation-Type Technology

The APS biosynthetic genes of this invention can be expressed in heterologous organisms for the purposes of their production at greater quantities than might be possible from their native hosts. A suitable host for heterologous expression is *E. coli* and techniques for gene expression in *E. coli* are well known. For example, the cloned APS genes can be expressed in *E. coli* using the expression vector pKK223 as described in example 11. The cloned genes can be fused in transcriptional fusion, so as to use the available ribosome binding site cognate to the heterologous gene. This approach facilitates the expression of operons which encode more than one open reading frame as translation of the individual ORFs will thus be dependent on their cognate ribosome binding site signals. Alternatively APS genes can be fused to the vector's ATG (*e.g.* as an *NcoI* fusion) so as to use the *E. coli* ribosome binding site. For multiple ORF expression in *E. coli* (*e.g.* in the case of operons with multiple ORFs) this type of construct would require a separate promoter to be fused to each ORF. It is possible, however, to fuse the first ATG of the APS operon to the *E. coli* ribosome binding site while requiring the other ORFs to utilize their cognate ribosome binding sites. These types of construction for the overexpression of genes in *E. coli* are well known in the art. Suitable bacterial promoters include the *lac* promoter, the *tac* (*trp/lac*) promoter, and the Pλ promoter from bacteriophage λ. Suitable commercially available vectors include, for example, pKK223-3, pKK233-2, pDR540, pDR720, pYEJ001 and pPL-Lambda (from Pharmacia, Piscataway, NJ).

Similarly, gram positive bacteria, notably *Bacillus* species and particularly *Bacillus licheniformis*, are used in commercial scale production of heterologous proteins and can be adapted to the expression of APS biosynthetic genes (*e.g.* Quax *et al.*, *In*: Industrial Microorganisms: Basic and Applied Molecular Genetics, *Eds.:* Baltz *et al.*, American Society for Microbiology, Washington (1993)). Regulatory signals from a highly expressed *Bacillus* gene (*e.g.* amylase promoter, Quax *et al.*, *supra*) are used to generate transcriptional fusions with the APS biosynthetic genes.

In some instances, high level expression of bacterial genes has been achieved using yeast systems, such as the methylotrophic yeast *Pichia pastoris* (Sreekrishna, *In*: Industrial microorganisms: basic and applied molecular genetics, Baltz, Hegeman, and Skatrud *eds.*, American Society for Microbiology, Washington (1993)). The APS gene(s) of interest are positioned behind 5' regulatory sequences of the *Pichia* alcohol oxidase gene in vectors such as pHIL-D1 and pHIL-D2 (Sreekrishna, *supra*). Such vectors are used to transform *Pichia* and introduce the heterologous DNA into the yeast genome. Likewise, the yeast *Saccharomyces cerevisiae* has been used to express heterologous bacterial genes (*e.g.* Dequin & Barre, Biotechnology 12:173-177 (1994)). The yeast *Kluyveromyces lactis* is also a suitable host for heterologous gene expression (*e.g.* van den Berg *et al.*, Biotechnology 8:135-139 (1990)).

Overexpression of APS genes in organisms such as *E. coli, Bacillus* and yeast, which are known for their rapid growth and multiplication, will enable fermentation-production of larger quantities of APSs. The choice of organism may be restricted by the possible susceptibility of the organism to the APS being overproduced; however, the likely susceptibility can be determined by the procedures outlined in Section J. The APSs can be isolated and purified from such cultures (see "G") for use in the control of microorganisms such as fungi and bacteria.

I. Expression of Antibiotic Biosynthetic Genes in Microbial Hosts for Biocontrol Purposes

The cloned APS biosynthetic genes of this invention can be utilized to increase the efficacy of biocontrol strains of various microorganisms. One possibility is the transfer of the genes for a particular APS back into its native host under stronger transcriptional regulation to cause the production of larger quantities of the APS. Another possibility is the transfer of genes to a heterologous host, causing production in the heterologous host of an APS not normally produced by that host.

Microorganisms which are suitable for the heterologous overexpression of APS genes are all microorganisms which are capable of colonizing plants or the rhizosphere. As such they will be brought into contact with phytopathogenic fungi causing an inhibition of their growth. These include gram-negative microorganisms such as *Pseudomonas, Enterobacter* and

Serratia, the gram-positive microorganism *Bacillus* and *Streptomyces spp.* and the fungi *Trichoderma* and *Gliocladium*. Particularly preferred heterologous hosts are *Pseudomonas fluorescens, Pseudomonas putida, Pseudomonas cepacia, Pseudomonas aureofaciens, Pseudomonas aurantiaca, Enterobacter cloacae, Serratia marscesens, Bacillus subtilis, Bacillus cereus, Trichoderma viride, Trichoderma harzianum* and *Gliocladium virens.*

Example 27: **Expression of APS Biosynthetic Genes in *E coli* and Other Gram-Negative Bacteria**

Many genes have been expressed in gram-negative bacteria in a heterologous manner. Example 11 describes the expression of genes for pyrrolnitrin biosynthesis in *E. coli* using the expression vector pKK223-3 (Pharmacia catalogue # 27-4935-01). This vector has a strong *tac* promoter (Brosius, J. *et al., Proc. Natl. Acad. Sci. USA 81)* regulated by the *lac* repressor and induced by IPTG. A number of other expression systems have been developed for use in *E. coli* and some are detailed in Examples 14-17 above. The thermoinducible expression vector pP_L (Pharmacia #27-4946-01) uses a tightly regulated bacteriophage λ promoter which allows for high level expression of proteins. The *lac* promoter provides another means of expression but the promoter is not expressed at such high levels as the *tac* promoter. With the addition of broad host range replicons to some of these expression system vectors, production of antifungal compounds in closely related gram negative-bacteria such as *Pseudomonas, Enterobacter, Serratia* and *Erwinia* is possible. For example, pLRKD211 (Kaiser & Kroos, Proc. Natl. Acad. Sci. USA 81: 5816-5820 (1984)) contains the broad host range replicon *ori T* which allows replication in many gram-negative bacteria.

In *E coli*, induction by IPTG is required for expression of the *tac* (*i.e. trp-lac*) promoter. When this same promoter (*e.g.* on wide-host range plasmid pLRKD211) is introduced into *Pseudomonas* it is constitutively active without induction by IPTG. This *trp-lac* promoter can be placed in front of any gene or operon of interest for expression in *Pseudomonas* or any other closely related bacterium for the purposes of the constitutive expression of such a gene. If the operon of interest contains the information for the biosynthesis of an APS, then an otherwise biocontrol-minus strain of a gram-negative bacterium may be able to protect plants against a variety of fungal diseases. Thus, genes for antifungal compounds can therefore be placed behind a strong constitutive promoter, transferred to a bacterium that

normally does not produce antifungal products and which has plant or rhizosphere colonizing properties turning these organisms into effective biocontrol strains. Other possible promoters can be used for the constitutive expression of APS genes in gram-negative bacteria. These include, for example, the promoter from the *Pseudomonas* regulatory genes *gafA* and *lemA* (WO 94/01561) and the *Pseudomonas savastanoi* IAA operon promoter (Gaffney *et al., J. Bacteriol. 172:* 5593-5601 (1990).

The synthetic Prn operon with the tac promoter as described in example 11a was inserted into two broad host range vectors that replicate in a wide range of Gram negative bacteria. The first vector, pRK290 (Ditta *et al* 1980. PNAS 77(12) pp. 7347-7351), is a low copy number plasmid and the second vector, pBBR1MCS (Kovach *et al* 1994, Biotechniques 16(5):800-802), a medium copy number plasmid. Constructs of both vectors containing the Prn genes were introduced into a number of Gram negative bacterial strains and assayed for production of Pyrrolnitrin by TLC and HPLC. A number of strains were shown to heterologously produce Pyrrolnitirn. These include *E.coli, Pseudomonas sp.* (MOCG133, MOCG380, MOCG382, BL897, BL1889, BL2595) and *Enterobacter taylorae* (MOCG206).

Example 28: **Expression of APS Biosynthetic Genes in Gram-Positive Bacteria**
Heterologous expression of genes encoding APS genes in gram-positive bacteria is another means of producing new biocontrol strains. Expression systems for *Bacillus* and *Streptomyces* are the best characterized. The promoter for the erythromycin resistance gene (*ermR*) from *Streptococcus pneumoniae* has been shown to be active in gram-positive aerobes and anaerobes and also in *E.coli* (Trieu-Cuot *et al.*, Nucl Acids Res 18: 3660 (1990)). A further antibiotic resistance promoter from the thiostreptone gene has been used in *Streptomyces* cloning vectors (Bibb, Mol Gen Genet 199: 26-36 (1985)). The shuttle vector pHT3101 is also appropriate for expression in *Bacillus* (Lereclus, FEMS Microbiol Lett 60: 211-218 (1989)). By expressing an operon (such as the pyrrolnitrin operon) or individual APS encoding genes under control of the *ermR* or other promoters it will be possible to convert soil bacilli into strains able to protect plants against microbial diseases. A significant advantage of this approach is that many gram-positive bacteria produce spores which can be used in formulations that produce biocontrol products with a longer shelf life. *Bacillus* and *Streptomyces* species are aggressive colonizers of soils. In fact both produce secondary metabolites including antibiotics active against a broad range of

organisms and the addition of heterologous antifungal genes including (including those encoding pyrrolnitrin, soraphen, phenazine or cyclic peptides) to gram-positive bacteria may make these organisms even better biocontrol strains.

Example 29: **Expression of APS Biosynthetic Genes in Fungi**

Trichoderma harzianum and *Gliocladium virens* have been shown to provide varying levels of biocontrol in the field (US 5,165,928 and US 4,996,157, both to Cornell Research Foundation). The successful use of these biocontrol agents will be greatly enhanced by the development of improved strains by the introduction of genes for APSs. This could be accomplished by a number of ways which are well known in the art. One is protoplast mediated transformation of the fungus by PEG or electroporation-mediated techniques. Alternatively, particle bombardment can be used to transform protoplasts or other fungal cells with the ability to develop into regenerated mature structures. The vector pAN7-1, originally developed for *Aspergillus* transformation and now used widely for fungal transformation (Curragh *et al., Mycol. Res. 97(3):* 313-317 (1992); Tooley *et al., Curr. Genet. 21:* 55-60 (1992); Punt *et al.*, Gene 56: 117-124 (1987)) is engineered to contain the pyrrolnitrin operon, or any other genes for APS biosynthesis. This plasmid contains the *E. coli* the hygromycin B resistance gene flanked by the *Aspergillus nidulans gpd* promoter and the *trpC* terminator (Punt *et al.*, Gene 56: 117-124 (1987)).

J. *In Vitro* Activity of Anti-phytopathogenic Substances Against Plant Pathogens

Example 30: **Bioassay Procedures for the Detection of Antifungal Activity**

Inhibition of fungal growth by a potential antifungal agent can be determined in a number of assay formats. Macroscopic methods which are commonly used include the agar diffusion assay (Dhingra & Sinclair, Basic Plant Pathology Methods, CRC Press, Boca Raton, FLA (1985)) and assays in liquid media (Broekaert *et al.*, FEMS Microbiol. Lett. 69: 55-60.(1990)). Both types of assay are performed with either fungal spores or mycelia as inocula. The maintenance of fungal stocks is in accordance with standard mycological procedures. Spores for bioassay are harvested from a mature plate of a fungus by flushing the surface of the culture with sterile water or buffer. A suspension of mycelia is prepared by placing fungus from a plate in a blender and homogenizing until the colony is dispersed. The homogenate is filtered through several layers of cheesecloth so that larger particles are

excluded. The suspension which passes through the cheesecloth is washed by centrifugation and replacing the supernatant with fresh buffer. The concentration of the mycelial suspension is adjusted empirically, by testing the suspension in the bioassay to be used.

Agar diffusion assays may be performed by suspending spores or mycelial fragments in a solid test medium, and applying the antifungal agent at a point source, from which it diffuses. This may be done by adding spores or mycelia to melted fungal growth medium, then pouring the mixture into a sterile dish and allowing it to gel. Sterile filters are placed on the surface of the medium, and solutions of antifungal agents are spotted onto the filters. After the liquid has been absorbed by the filter, the plates are incubated at the appropriate temperature, usually for 1-2 days. Growth inhibition is indicated by the presence of zones around filters in which spores have not germinated, or in which mycelia have not grown. The antifungal potency of the agent, denoted as the minimal effective dose, may be quantified by spotting serial dilutions of the agent onto filters, and determining the lowest dose which gives an observable inhibition zone. Another agar diffusion assay can be performed by cutting wells into solidified fungal growth medium and placing solutions of antifungal agents into them. The plate is inoculated at a point equidistant from all the wells, usually at the center of the plate, with either a small aliquot of spore or mycelial suspension or a mycelial plug cut directly from a stock culture plate of the fungus. The plate is incubated for several days until the growing mycelia approach the wells, then it is observed for signs of growth inhibition. Inhibition is indicated by the deformation of the roughly circular form which the fungal colony normally assumes as it grows. Specifically, if the mycelial front appears flattened or even concave relative to the uninhibited sections of the plate, growth inhibition has occurred. A minimal effective concentration may be determined by testing diluted solutions of the agent to find the lowest at which an effect can be detected.

Bioassays in liquid media are conducted using suspensions of spores or mycelia which are incubated in liquid fungal growth media instead of solid media. The fungal inocula, medium, and antifungal agent are mixed in wells of a 96-well microtiter plate, and the growth of the fungus is followed by measuring the turbidity of the culture spectrophotometrically. Increases in turbidity correlate with increases in biomass, and are a measure of fungal

growth. Growth inhibition is determined by comparing the growth of the fungus in the presence of the antifungal agent with growth in its absence. By testing diluted solutions of antifungal inhibitor, a minimal inhibitory concentration or an EC_{50} may be determined.

Example 31: **Bioassay Procedures for the Detection of Antibacterial Activity**

A number of bioassays may be employed to determine the antibacterial activity of an unknown compound. The inhibition of bacterial growth in solid media may be assessed by dispersing an inoculum of the bacterial culture in melted medium and spreading the suspension evenly in the bottom of a sterile Petri dish. After the medium has gelled, sterile filter disks are placed on the surface, and aliquots of the test material are spotted onto them. The plate is incubated overnight at an appropriate temperature, and growth inhibition is observed as an area around a filter in which the bacteria have not grown, or in which the growth is reduced compared to the surrounding areas. Pure compounds may be characterized by the determination of a minimal effective dose, the smallest amount of material which gives a zone of inhibited growth. In liquid media, two other methods may be employed. The growth of a culture may be monitored by measuring the optical density of the culture, in actuality the scattering of incident light. Equal inocula are seeded into equal culture volumes, with one culture containing a known amount of a potential antibacterial agent. After incubation at an appropriate temperature, and with appropriate aeration as required by the bacterium being tested, the optical densities of the cultures are compared. A suitable wavelength for the comparison is 600 nm. The antibacterial agent may be characterized by the determination of a minimal effective dose, the smallest amount of material which produces a reduction in the density of the culture, or by determining an EC_{50}, the concentration at which the growth of the test culture is half that of the control. The bioassays described above do not differentiate between bacteriostatic and bacteriocidal effects. Another assay can be performed which will determine the bacteriocidal activity of the agent. This assay is carried out by incubating the bacteria and the active agent together in liquid medium for an amount of time and under conditions which are sufficient for the agent to exert its effect. After this incubation is completed, the bacteria may be either washed by centrifugation and resuspension, or diluted by the addition of fresh medium. In either case, the concentration of the antibacterial agent is reduced to a point at which it is no longer expected to have significant activity. The bacteria are plated and spread on solid medium and the plates are incubated overnight at an appropriate

temperature for growth. The number of colonies which arise on the plates are counted, and the number which appeared from the mixture which contained the antibacterial agent is compared with the number which arose from the mixture which contained no antibacterial agent. The reduction in colony-forming units is a measure of the bacteriocidal activity of the agent. The bacteriocidal activity may be quantified as a minimal effective dose, or as an EC_{50}, as described above. Bacteria which are used in assays such as these include species of *Agrobacterium, Erwinia, Clavibacter, Xanthomonas,* and *Pseudomonas.*

Example 32: **Antipathogenic Activity Determination of APSs**

APSs are assayed using the procedures of examples 30 and 31 above to identify the range of fungi and bacteria against which they are active. The APS can be isolated from the cells and culture medium of the host organism normally producing it, or can alternatively be isolated from a heterologous host which has been engineered to produce the APS. A further possibility is the chemical synthesis of APS compounds of known chemical structure, or derivatives thereof.

Example 33: **Antimicriobial Activity Determination of Pyrrolnitrin**

a) The anti-phytopathogenic activity of a fluorinated 3-cyano-derivative of pyrrolnitrin (designated CGA173506) was observed against the maize fungal phytopathgens *Diplodia maydis, Colletotrichum graminicola,* and *Gibberella zeae-maydis.* Spores of the fungi were harvested and suspended in water. Approximately 1000 spores were inoculated into potato dextrose broth and either CGA173506 or water in a total volume of 100 microliters in the wells of 96-well microtiter plates suitable for a plate reader. The compound CGA173506 was obtained as a 50% wettable powder, and a stock suspension was made up at a concentration of 10 mg/ml in sterile water. This stock suspension was diluted with sterile water to provide the 173506 used in the tests. After the spores, medium, and 173506 were mixed, the turbidity in the wells was measured by reading the absorbance at 600 nm in a plate reader. This reading was taken as the background turbidity, and was subtracted from readings taken at later times. After 46 hours of incubation, the presence of 1 microgram/ml of 173506 was determined to reduce the growth of *Diplodia maydis* by 64%, and after 120 hours, the same concentration of 173506 inhibited the growth of *Colletotrichum graminicola* by 50%. After 40 hours of incubation, the presence of 0.5 microgram/ml of 173506 gave 100% inhibition of *Gibberella zeae-maydis.*

b) Pyrrolnitrin was tested for its effect on the growth of various maize fungal pathogens and inibited growth of *Bipolaris maydis*, *Colletotrichum graminicola*, *Diplodia maydis*, *Fusarium moniliforme*, *Gibberella zeae* and *Rhizoctania solani*.

To determine growth

To determine growth inhibition autoclaved filter discs (0.25 inch diameter from Schleicher and Schuell) were placed near the perimeter of PDA (DIFCO) plates. Solutions were pipetted onto these filters. 2.5 micrograms pyrrolnitrin (25 microliter) were placed on one filter disc and 25 microliters 63% ethanol were placed on the other disc. Fungal plugs were taken from stock plates and placed in the center of the PDA plates. Each fungus was inoculated onto one plate. the fungus was allowed to grow and inhibition was scored at appropriate times. Inhibition of the fungi indicated above was visually detected.

K. Expression of Antibiotic Biosynthetic Genes in Transgenic Plants

Example 34: **Modification of Coding Sequences and Adjacent Sequences**

The cloned APS biosynthetic genes described in this application can be modified for expression in transgenic plant hosts. This is done with the aim of producing extractable quantities of APS from transgenic plants (*i.e.* for similar reasons to those described in Section E above), or alternatively the aim of such expression can be the accumulation of APS in plant tissue for the provision of pathogen protection on host plants. A host plant expressing genes for the biosynthesis of an APS and which produces the APS in its cells will have enhanced resistance to phytopathogen attack and will be thus better equipped to withstand crop losses associated with such attack.

The transgenic expression in plants of genes derived from microbial sources may require the modification of those genes to achieve and optimize their expression in plants. In particular, bacterial ORFs which encode separate enzymes but which are encoded by the same transcript in the native microbe are best expressed in plants on separate transcripts. To achieve this, each microbial ORF is isolated individually and cloned within a cassette which provides a plant promoter sequence at the 5' end of the ORF and a plant transcriptional terminator at the 3' end of the ORF. The isolated ORF sequence preferably includes the initiating ATG codon and the terminating STOP codon but may include

additional sequence beyond the initiating ATG and the STOP codon. In addition, the ORF may be truncated, but still retain the required activity; for particularly long ORFs, truncated versions which retain activity may be preferable for expression in transgenic organisms. By "plant promoter" and "plant transcriptional terminator" it is intended to mean promoters and transcriptional terminators which operate within plant cells. This includes promoters and transcription terminators which may be derived from non-plant sources such as viruses (an example is the Cauliflower Mosaic Virus).

In some cases, modification to the ORF coding sequences and adjacent sequence will not be required. It is sufficient to isolate a fragment containing the ORF of interest and to insert it downstream of a plant promoter. For example, Gaffney *et al.* (Science 261: 754-756 (1993)) have expressed the *Pseudomonas nahG* gene in transgenic plants under the control of the CaMV 35S promoter and the CaMV *tml* terminator successfully without modification of the coding sequence and with 56 bp of the *Pseudomonas* gene upstream of the ATG still attached, and 165 bp downstream of the STOP codon still attached to the *nahG* ORF. Preferably as little adjacent microbial sequence should be left attached upstream of the ATG and downstream of the STOP codon. In practice, such construction may depend on the availability of restriction sites.

In other cases, the expression of genes derived from microbial sources may provide problems in expression. These problems have been well characterized in the art and are particularly common with genes derived from certain sources such as *Bacillus.* These problems may apply to the APS biosynthetic genes of this invention and the modification of these genes can be undertaken using techniques now well known in the art. The following problems may be encountered:

(1) Codon Usage. The preferred codon usage in plants differs from the preferred codon usage in certain microorganisms. Comparison of the usage of codons within a cloned microbial ORF to usage in plant genes (and in particular genes from the target plant) will enable an identification of the codons within the ORF which should preferably be changed. Typically plant evolution has tended towards a strong preference of the nucleotides C and G in the third base position of monocotyledons, whereas dicotyledons often use the nucleotides A or T at this position. By modifying a gene to incorporate preferred codon

usage for a particular target transgenic species, many of the problems described below for GC/AT content and illegitimate splicing will be overcome.

(2) GC/AT Content. Plant genes typically have a GC content of more than 35%. ORF sequences which are rich in A and T nucleotides can cause several problems in plants. Firstly, motifs of ATTTA are believed to cause destabilization of messages and are found at the 3' end of many short-lived mRNAs. Secondly, the occurrence of polyadenylation signals such as AATAAA at inappropriate positions within the message is believed to cause premature truncation of transcription. In addition, monocotyledons may recognize AT-rich sequences as splice sites (see below).

(3) Sequences Adjacent to the Initiating Methionine. Plants differ from microorganisms in that their messages do not possess a defined ribosome binding site. Rather, it is believed that ribosomes attach to the 5' end of the message and scan for the first available ATG at which to start translation. Nevertheless, it is believed that there is a preference for certain nucleotides adjacent to the ATG and that expression of microbial genes can be enhanced by the inclusion of a eukaryotic consensus translation initiator at the ATG. Clontech (1993/1994 catalog, page 210) have suggested the sequence GTCGACCATGGTC (SEQ ID NO:7) as a consensus translation initiator for the expression of the *E. coli uidA* gene in plants. Further, Joshi (NAR 15: 6643-6653 (1987)) has compared many plant sequences adjacent to the ATG and suggests the consensus TAAACAATGGCT (SEQ ID NO:8). In situations where difficulties are encountered in the expression of microbial ORFs in plants, inclusion of one of these sequences at the initiating ATG may improve translation. In such cases the last three nucleotides of the consensus may not be appropriate for inclusion in the modified sequence due to their modification of the second AA residue. Preferred sequences adjacent to the initiating methionine may differ between different plant species. A survey of 14 maize genes located in the GenBank database provided the following results:

Position Before the Initiating ATG in 14 Maize Genes:

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
C	3	8	4	6	2	5	6	0	10	7
T	3	0	3	4	3	2	1	1	1	0
A	2	3	1	4	3	2	3	7	2	3
G	6	3	6	0	6	5	4	6	1	5

This analysis can be done for the desired plant species into which APS genes are being incorporated, and the sequence adjacent to the ATG modified to incorporate the preferred nucleotides.

(4) Removal of Illegitimate Splice Sites. Genes cloned from non-plant sources and not optimized for expression in plants may also contain motifs which may be recognized in plants as 5' or 3' splice sites, and be cleaved, thus generating truncated or deleted messages.

Techniques for the modification of coding sequences and adjacent sequences are well known in the art. In cases where the initial expression of a microbial ORF is low and it is deemed appropriate to make alterations to the sequence as described above, then the construction of synthetic genes can be accomplished according to methods well known in the art. These are, for example, described in the published patent disclosures EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol) and WO 93/07278 (to Ciba-Geigy). In most cases it is preferable to assay the expression of gene constructions using transient assay protocols (which are well known in the art) prior to their transfer to transgenic plants.

Example 35: **Construction of Plant Transformation Vectors**

Numerous transformation vectors are available for plant transformation, and the genes of this invention can be used in conjunction with any such vectors. The selection of vector for use will depend upon the preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the *nptII* gene

which confers resistance to kanamycin and related antibiotics (Messing & Vierra, Gene 19: 259-268 (1982); Bevan *et al.*, Nature 304:184-187 (1983)), the *bar* gene which confers resistance to the herbicide phosphinothricin (White *et al.*, Nucl Acids Res 18: 1062 (1990), Spencer *et al.* Theor Appl Genet 79: 625-631(1990)), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, Mol Cell Biol 4: 2929-2931), and the *dhfr* gene, which confers resistance to methotrexate (Bourouis *et al.*, EMBO J. 2(7): 1099-1104 (1983)).

(1) Construction of Vectors Suitable for *Agrobacterium* Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, Nucl. Acids Res. (1984)). Below the construction of two typical vectors is described.

Construction of pCIB200 and pCIB2001

The binary vectors pCIB200 and pCIB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and was constructed in the following manner. pTJS75kan was created by *NarI* digestion of pTJS75 (Schmidhauser & Helinski, J Bacteriol. 164: 446-455 (1985)) allowing excision of the tetracycline-resistance gene, followed by insertion of an *AccI* fragment from pUC4K carrying an NPTII (Messing & Vierra, Gene 19: 259-268 (1982); Bevan *et al.*, Nature 304: 184-187 (1983); McBride *et al.*, Plant Molecular Biology 14: 266-276 (1990)). *XhoI* linkers were ligated to the *EcoRV* fragment of pCIB7 which contains the left and right T-DNA borders, a plant selectable *nos/nptII* chimeric gene and the pUC polylinker (Rothstein *et al.*, Gene 53: 153-161 (1987)), and the *XhoI*-digested fragment was cloned into *SalI*-digested pTJS75kan to create pCIB200 (see also EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: *EcoRI, SstI, KpnI, BglII, XbaI,* and *SalI*. *pCIB2001* is a derivative of pCIB200 which was created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are *EcoRI, SstI, KpnI, BglII, XbaI, SalI, MluI, BclI, AvrII, ApaI, HpaI,* and *StuI*. pCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived *trfA* function for mobilization between *E. coli* and other

hosts, and the *OriT* and *OriV* functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

Construction of pCIB10 and Hygromycin Selection Derivatives thereof

The binary vector pCIB10 contains a gene encoding kanamycin resistance for selection in plants, T-DNA right and left border sequences and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein *et al.* (Gene 53: 153-161 (1987)). Various derivatives of pCIB10 have been constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz *et al.* (Gene 25: 179-188 (1983)). These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

(2) Construction of Vectors Suitable for non-*Agrobacterium* Transformation.

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques which do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (*e.g.* PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Below, the construction of some typical vectors is described.

Construction of pCIB3064

pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide basta (or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites were mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites *SspI* and *PvuII*. The new restriction sites were 96 and 37 bp away from the unique *SalI* site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 was designated pCIB3025.

The GUS gene was then excised from pCIB3025 by digestion with *SalI* and *SacI*, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 was obtained from the John Innes Centre, Norwich and the a 400 bp *SmaI* fragment containing the *bar* gene from *Streptomyces viridochromogenes* was excised and inserted into the *HpaI* site of pCIB3060 (Thompson *et al.* EMBO J 6: 2519-2523 (1987)). This generated pCIB3064 which comprises the *bar* gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene for ampicillin resistance (for selection in *E. coli*) and a polylinker with the unique sites *SphI, PstI, HindIII,* and *BamHI*. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

Construction of pSOG19 and pSOG35

pSOG35 is a transformation vector which utilizes the *E. coli* gene dihydrofolate reductase (DHFR) as a selectable marker conferring resistance to methotrexate. PCR was used to amplify the 35S promoter (~800 bp), intron 6 from the maize Adh1 gene (~550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250 bp fragment encoding the *E. coli* dihydrofolate reductase type II gene was also amplified by PCR and these two PCR fragments were assembled with a *SacI-PstI* fragment from pBI221 (Clontech) which comprised the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generated pSOG19 which contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generated the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII, SphI, PstI* and *EcoRI* sites available for the cloning of foreign sequences.

Example 36: **Requirements for Construction of Plant Expression Cassettes**

Gene sequences intended for expression in transgenic plants are firstly assembled in expression cassettes behind a suitable promoter and upstream of a suitable transcription terminator. These expression cassettes can then be easily transferred to the plant transformation vectors described above in example 2-6.

Promoter Selection

The selection of promoter used in expression cassettes will determine the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters will express transgenes in specific cell types (such as leaf epidermal cells, meosphyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example) and this selection will reflect the desired location of biosynthesis of the APS. Alternatively, the selected promoter may drive expression of the gene under a light-induced or other temporally regulated promoter. A further alternative is that the selected promoter be chemically regulated. This would provide the possibility of inducing the induction of the APS only when desired and caused by treatment with a chemical inducer.

Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for the termination of transcription beyond the transgene and its correct polyadenylation. Appropriate transcriptional terminators and those which are known to function in plants and include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator, the pea *rbcS* E9 terminator. These can be used in both monocoylyedons and dicotyledons.

Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *Adh1* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, Genes Develep 1: 1183-1200 (1987)). In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression (Callis *et al.*, *supra*). Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV, the "Ω-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (*e.g.* Gallie *et al.* Nucl. Acids Res. 15: 8693-8711 (1987); Skuzeski *et al.* Plant Molec. Biol. 15; 65-79 (1990))

Targeting of the Gene Product Within the Cell

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the aminoterminal end of various proteins and which is cleaved during chloroplast import yielding the mature protein (*e.g.* Comai *et al.* J. Biol. Chem. 263: 15104-15109 (1988)). These signal sequences can be fused to heterologous gene products to effect the import of heterologous products into the chloroplast (van den Broeck *et al.* Nature 313: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the EPSP synthase enzyme, the GS2 protein and many other proteins which are known to be chloroplast localized.

Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (*e.g.* Unger *et al.* Plant Molec. Biol. 13: 411-418 (1989)). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting to cellular protein bodies has been described by Rogers *et al.* (Proc. Natl. Acad. Sci. USA 82: 6512-6516 (1985)).

In addition sequences have been characterized which cause the targeting of gene products to other cell compartments. Aminoterminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, Plant Cell 2: 769-783 (1990)). Additionally, aminoterminal sequences in conjunction with

carboxyterminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.* Plant Molec. Biol. 14: 357-368 (1990)).

By the fusion of the appropriate targeting sequences described above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the aminoterminal ATG of the transgene. The signal sequence selected should include the known cleavage site and the fusion constructed should take into account any amino acids after the cleavage site which are required for cleavage. In some cases this requirement may be fulfilled by the addition of a small number of amino acids between the cleavage site and the transgene ATG or alternatively replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by *in vitro* translation of *in vitro* transcribed constructions followed by *in vitro* chloroplast uptake using techniques described by (Bartlett *et al.* In: Edelmann *et al.* (Eds.) Methods in Chloroplast Molecular Biology, Elsevier. pp 1081-1091 (1982); Wasmann *et al.* Mol. Gen. Genet. 205: 446-453 (1986)). These construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes. The choice of targeting which may be required for APS biosynthetic genes will depend on the cellular localization of the precursor required as the starting point for a given pathway. This will usually be cytosolic or chloroplastic, although it may is some cases be mitochondrial or peroxisomal. The gene products of APS biosynthetic genes will not normally require targeting to the ER, the apoplast or the vacuole.

The above described mechanisms for cellular targeting can be utilized not only in conjunction with their cognate promoters, but also in conjunction with heterologous promoters so as to effect a specific cell targeting goal under the transcriptional regulation of a promoter which has an expression pattern different to that of the promoter from which the targeting signal derives.

Example 37: **Examples of Expression Cassette Construction**

The present invention encompasses the expression of genes encoding APSs under the regulation of any promoter which is expressible in plants, regardless of the origin of the promoter.

Furthermore, the invention encompasses the use of any plant-expressible promoter in conjunction with any further sequences required or selected for the expression of the APS gene. Such sequences include, but are not restricted to, transcriptional terminators, extraneous sequences to enhance expression (such as introns (*e.g. Adh* intron 1), viral sequences (*e. g.* TMV-Ω)), and sequences intended for the targeting of the gene product to specific organelles and cell compartments.

Constitutive Expression: the CaMV 35S Promoter

Construction of the plasmid pCGN1761 is described in the published patent application EP 0 392 225 (example 23). pCGN1761 contains the "double" 35S promoter and the *tml* transcriptional terminator with a unique *EcoRI* site between the promoter and the terminator and has a pUC-type backbone. A derivative of pCGN1761 was constructed which has a modified polylinker which includes *NotI* and *XhoI* sites in addition to the existing *EcoRI* site. This derivative was designated pCGN1761ENX. pCGN1761ENX is useful for the cloning of cDNA sequences or gene sequences (including microbial ORF sequences) within its polylinker for the purposes of their expression under the control of the 35S promoter in transgenic plants. The entire 35S promoter-gene sequence-*tml* terminator cassette of such a construction can be excised by *HindIII, SphI, SalI,* and *XbaI* sites 5' to the promoter and *XbaI, BamHI* and *BglI* sites 3' to the terminator for transfer to transformation vectors such as those described above in example 35. Furthermore, the double 35S promoter fragment can be removed by 5' excision with *HindIII, SphI, SalI, XbaI,* or *PstI,* and 3' excision with any of the polylinker restriction sites (*EcoRI, NotI* or *XhoI*) for replacement with another promoter.

Modification of pCGN1761ENX by Optimization of the Translational Initiation Site

For any of the constructions described in this section, modifications around the cloning sites can be made by the introduction of sequences which may enhance translation. This is particularly useful when genes derived from microorganisms are to be introduced into plant

expression cassettes as these genes may not contain sequences adjacent to their initiating methionine which may be suitable for the initiation of translation in plants. In cases where genes derived from microorganisms are to be cloned into plant expression cassettes at their ATG it may be useful to modify the site of their insertion to optimize their expression. Modification of pCGN1761ENX is described by way of example to incorporate one of several optimized sequences for plant expression (*e.g.* Joshi, NAR 15: 6643-6653 (1987)).

pCGN1761ENX is cleaved with *SphI*, treated with T4 DNA polymerase and religated, thus destroying the *SphI* site located 5' to the double 35S promoter. This generates vector pCGN1761ENX/Sph-. pCGN1761ENX/Sph- is cleaved with *EcoRI*, and ligated to an annealed molecular adaptor of the sequence 5'-AATTCTAAAGCATGCCGATCGG-3'(SEQ ID NO:9)/5'-AATTCCGATCGGCATGCTTTA-3' (SEQ ID NO:10). This generates the vector pCGNSENX which incorporates the *quasi*-optimized plant translational initiation sequence TAAA-C adjacent to the ATG which is itself part of an *SphI* site which is suitable for cloning heterologous genes at their initiating methionine. Downstream of the *SphI* site, the *EcoRI*, *NotI*, and *XhoI* sites are retained.

An alternative vector is constructed which utilizes an *NcoI* site at the initiating ATG. This vector, designated pCGN1761NENX is made by inserting an annealed molecular adaptor of the sequence 5'-AATTCTAAACCATGGCGATCGG-3' (SEQ ID NO:11) / 5'AATTCCGATCGCCATGGTTTA-3' (SEQ ID NO:12) at the pCGN1761ENX *EcoRI* site (Sequence ID's 14 and 15). Thus, the vector includes the *quasi*-optimized sequence TAAACC adjacent to the initiating ATG which is within the *NcoI* site. Downstream sites are *EcoRI*, *NotI*, and *XhoI*. Prior to this manipulation, however, the two *NcoI* sites in the pCGN1761ENX vector (at upstream positions of the 5' 35S promoter unit) are destroyed using similar techniques to those described above for *SphI* or alternatively using "inside-outside" PCR (Innes *et al.* PCR Protocols: A guide to methods and applications. Academic Press, New York (1990); see Example 41). This manipulation can be assayed for any possible detrimental effect on expression by insertion of any plant cDNA or reporter gene sequence into the cloning site followed by routine expression analysis in plants.

Expression under a Chemically Regulatable Promoter

This section describes the replacement of the double 35S promoter in pCGN1761ENX with any promoter of choice; by way of example the chemically regulated PR-1a promoter is described. The promoter of choice is preferably excised from its source by restriction enzymes, but can alternatively be PCR-amplified using primers which carry appropriate terminal restriction sites. Should PCR-amplification be undertaken, then the promoter should be resequenced to check for amplification errors after the cloning of the amplified promoter in the target vector. The chemically regulatable tobacco PR-1a promoter is cleaved from plasmid pCIB1004 (see EP 0 332 104, example 21 for construction) and transferred to plasmid pCGN1761ENX. pCIB1004 is cleaved with *NcoI* and the resultant 3' overhang of the linearized fragment is rendered blunt by treatment with T4 DNA polymerase. The fragment is then cleaved with *HindIII* and the resultant PR-1a promoter containing fragment is gel purified and cloned into pCGN1761ENX from which the double 35S promoter has been removed. This is done by cleavage with *XhoI* and blunting with T4 polymerase, followed by cleavage with *HindIII* and isolation of the larger vector-terminator containing fragment into which the pCIB1004 promoter fragment is cloned. This generates a pCGN1761ENX derivative with the PR-1a promoter and the *tml* terminator and an intervening polylinker with unique *EcoRI* and *NotI* sites. Selected APS genes can be inserted into this vector, and the fusion products (*i.e.* promoter-gene-terminator) can subsequently be transferred to any selected transformation vector, including those described in this application.

Constitutive Expression: the Actin Promoter

Several isoforms of actin are known to be expressed in most cell types and consequently the actin promoter is a good choice for a constitutive promoter. In particular, the promoter from the rice *Act1* gene has been cloned and characterized (McElroy *et al.* Plant Cell 2: 163-171 (1990)). A 1.3 kb fragment of the promoter was found to contain all the regulatory elements required for expression in rice protoplasts. Furthermore, numerous expression vectors based on the *Act1* promoter have been constructed specifically for use in monocotyledons (McElroy *et al.* Mol. Gen. Genet. 231: 150-160 (1991)). These incorporate the *Act1*-intron 1, *Adh1* 5' flanking sequence and *Adh1*-intron 1 (from the maize alcohol dehydrogenase gene) and sequence from the CaMV 35S promoter. Vectors showing highest expression were fusions of 35S and the *Act1* intron or the *Act1* 5' flanking sequence

and the *Act1* intron. Optimization of sequences around the initiating ATG (of the GUS reporter gene) also enhanced expression. The promoter expression cassettes described by McElroy *et al.* (Mol. Gen. Genet. 231: 150-160 (1991)) can be easily modified for the expression of APS biosynthetic genes and are particularly suitable for use in monocotyledonous hosts. For example, promoter containing fragments can be removed from the McElroy constructions and used to replace the double 35S promoter in pCGN1761ENX, which is then available for the insertion of specific gene sequences. The fusion genes thus constructed can then be transferred to appropriate transformation vectors. In a separate report the rice *Act1* promoter with its first intron has also been found to direct high expression in cultured barley cells (Chibbar *et al.* Plant Cell Rep. 12: 506-509 (1993)).

Constitutive Expression: the Ubiquitin Promoter

Ubiquitin is another gene product known to accumulate in many call types and its promoter has been cloned from several species for use in transgenic plants (*e.g.* sunflower - Binet *et al.* Plant Science 79: 87-94 (1991), maize - Christensen *et al.* Plant Molec. Biol. 12: 619-632 (1989)). The maize ubiquitin promoter has been developed in transgenic monocot systems and its sequence and vectors constructed for monocot transformation are disclosed in the patent publication EP 0 342 926 (to Lubrizol). Further, Taylor *et al.* (Plant Cell Rep. 12: 491-495 (1993)) describe a vector (pAHC25) which comprises the maize ubiquitin promoter and first intron and its high activity in cell suspensions of numerous monocotyledons when introduced via microprojectile bombardment. The ubiquitin promoter is clearly suitable for the expression of APS biosynthetic genes in transgenic plants, especially monocotyledons. Suitable vectors are derivatives of pAHC25 or any of the transformation vectors described in this application, modified by the introduction of the appropriate ubiquitin promoter and/or intron sequences.

Root Specific Expression

A preferred pattern of expression for the APSs of the instant invention is root expression. Root expression is particularly useful for the control of soil-borne phytopathogens such as *Rhizoctonia* and *Pythium*. Expression of APSs only in root tissue would have the advantage of controlling root invading phytopathogens, without a concomitant accumulation of APS in leaf and flower tissue and seeds. A suitable root promoter is that described by de

Framond (FEBS 290: 103-106 (1991)) and also in the published patent application EP 0 452 269 (to Ciba-Geigy). This promoter is transferred to a suitable vector such as pCGN1761ENX for the insertion of an APS gene of interest and subsequent transfer of the entire promoter-gene-terminator cassette to a transformation vector of interest.

Wound Inducible Promoters

Wound-inducible promoters are particularly suitable for the expression of APS biosynthetic genes because they are typically active not just on wound induction, but also at the sites of phytopathogen infection. Numerous such promoters have been described (*e.g.* Xu *et al.* Plant Molec. Biol. 22: 573-588 (1993), Logemann *et al.* Plant Cell 1: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22: 783-792 (1993), Firek *et al.* Plant Molec. Biol. 22: 129-142 (1993), Warner *et al.* Plant J. 3: 191-201 (1993)) and all are suitable for use with the instant invention. Logemann *et al.* (*supra*) describe the 5' upstream sequences of the dicotyledonous potato *wun1* gene. Xu *et al.* (*supra*) show that a wound inducible promoter from the dicotyledon potato (*pin2*) is active in the monocotyledon rice. Further, Rohrmeier & Lehle (*supra*) describe the cloning of the maize *Wip1* cDNA which is wound induced and which can be used to isolated the cognate promoter using standard techniques. Similarly, Firek *et al.* (*supra*) and Warner *et al.* (*supra*) have described a wound induced gene from the monocotyledon *Asparagus officinalis* which is expressed at local wound and pathogen invasion sites. Using cloning techniques well known in the art, these promoters can be transferred to suitable vectors, fused to the APS biosynthetic genes of this invention, and used to express these genes at the sites of phytopathogen infection.

Pith Preferred Expression

Patent Application WO 93/07278 (to Ciba-Geigy) describes the isolation of the maize *trpA* gene which is preferentially expressed in pith cells. The gene sequence and promoter extending up to nucleotide -1726 from the start of transcription are presented. Using standard molecular biological techniques, this promoter or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a foreign gene in a pith-preferred manner. In fact fragments containing the pith-preferred promoter or parts thereof can be transferred to any vector and modified for utility in transgenic plants.

Pollen-Specific Expression

Patent Application WO 93/07278 (to Ciba-Geigy) further describes the isolation of the maize calcium-dependent protein kinase (CDPK) gene which is expressed in pollen cells. The gene sequence and promoter extend up to 1400 bp from the start of transcription. Using standard molecular biological techniques, this promoter or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a foreign gene in a pollen-specific manner. In fact fragments containing the pollen-specific promoter or parts thereof can be transferred to any vector and modified for utility in transgenic plants.

Leaf-Specific Expression

A maize gene encoding phosphoenol carboxylase (PEPC) has been described by Hudspeth & Grula (Plant Molec Biol 12: 579-589 (1989)). Using standard molecular biological techniques the promoter for this gene can be used to drive the expression of any gene in a leaf-specific manner in transgenic plants.

Expression with Chloroplast Targeting

Chen & Jagendorf (J. Biol. Chem. 268: 2363-2367 (1993) have described the successful use of a chloroplast transit peptide for import of a heterologous transgene. This peptide used is the transit peptide from the *rbcS* gene from *Nicotiana plumbaginifolia* (Poulsen *et al.* Mol. Gen. Genet. 205: 193-200 (1986)). Using the restriction enzymes *DraI* and *SphI*, or *Tsp509I* and *SphI* the DNA sequence encoding this transit peptide can be excised from plasmid prbcS-8B (Poulsen *et al. supra*) and manipulated for use with any of the constructions described above. The *DraI-SphI* fragment extends from -58 relative to the initiating *rbcS* ATG to, and including, the first amino acid (also a methionine) of the mature peptide immediately after the import cleavage site, whereas the *Tsp509I-SphI* fragment extends from -8 relative to the initiating *rbcS* ATG to, and including, the first amino acid of the mature peptide. Thus, these fragment can be appropriately inserted into the polylinker of any chosen expression cassette generating a transcriptional fusion to the untranslated leader of the chosen promoter (*e.g.* 35S, PR-1a, actin, ubiquitin *etc.*), whilst enabling the insertion of a required APS gene in correct fusion downstream of the transit peptide. Constructions of this kind are routine in the art. For example, whereas the *DraI* end is already blunt, the 5' *Tsp509I* site may be rendered blunt by T4 polymerase treatment, or

may alternatively be ligated to a linker or adaptor sequence to facilitate its fusion to the chosen promoter. The 3' *SphI* site may be maintained as such, or may alternatively be ligated to adaptor or linker sequences to facilitate its insertion into the chosen vector in such a way as to make available appropriate restriction sites for the subsequent insertion of a selected APS gene. Ideally the ATG of the *SphI* site is maintained and comprises the first ATG of the selected APS gene. Chen & Jagendorf (*supra*) provide consensus sequences for ideal cleavage for chloroplast import, and in each case a methionine is preferred at the first position of the mature protein. At subsequent positions there is more variation and the amino acid may not be so critical. In any case, fusion constructions can be assessed for efficiency of import *in vitro* using the methods described by Bartlett *et al.* (In: Edelmann *et al.* (Eds.) Methods in Chloroplast Molecular Biology, Elsevier. pp 1081-1091 (1982)) and Wasmann *et al.* (Mol. Gen. Genet. 205: 446-453 (1986)). Typically the best approach may be to generate fusions using the selected APS gene with no modifications at the aminoterminus, and only to incorporate modifications when it is apparent that such fusions are not chloroplast imported at high efficiency, in which case modifications may be made in accordance with the established literature (Chen & Jagendorf, *supra*; Wasman *et al., supra*; Ko & Ko, J. Biol. Chem. 267: 13910-13916 (1992)).

A preferred vector is constructed by transferring the *DraI-SphI* transit peptide encoding fragment from prbcS-8B to the cloning vector pCGN1761ENX/Sph-. This plasmid is cleaved with *EcoRI* and the termini rendered blunt by treatment with T4 DNA polymerase. Plasmid prbcS-8B is cleaved with *SphI* and ligated to an annealed molecular adaptor of the sequence 5'-CCAGCTGGAATTCCG-3' (SEQ ID NO:13)/5'-CGGAATTCCAGCTGGCATG-3' (SEQ ID NO:14). The resultant product is 5'-terminally phosphorylated by treatment with T4 kinase. Subsequent cleavage with *DraI* releases the transit peptide encoding fragment which is ligated into the blunt-end ex-*EcoRI* sites of the modified vector described above. Clones oriented with the 5' end of the insert adjacent to the 3' end of the 35S promoter are identified by sequencing. These clones carry a DNA fusion of the 35S leader sequence to the *rbcS-8A* promoter-transit peptide sequence extending from -58 relative to the *rbcS* ATG to the ATG of the mature protein, and including at that position a unique *SphI* site, and a newly created *EcoRI* site, as well as the existing *NotI* and *XhoI* sites of pCGN1761ENX. This new vector is designated pCGN1761/CT. DNA sequences are transferred to pCGN1761/CT in frame by amplification using PCR techniques and incorporation of an

SphI, NsphI, or *NlaIII* site at the amplified ATG, which following restriction enzyme cleavage with the appropriate enzyme is ligated into *SphI*-cleaved pCGN1761/CT. To facilitate construction, it may be required to change the second amino acid of the cloned gene, however, in almost all cases the use of PCR together with standard site directed mutagenesis will enable the construction of any desired sequence around the cleavage site and first methionine of the mature protein.

A further preferred vector is constructed by replacing the double 35S promoter of pCGN1761ENX with the *BamHI-SphI* fragment of prbcS-8A which contains the full-length light regulated *rbcS-8A* promoter from nucleotide -1038 (relative to the transcriptional start site) up to the first methionine of the mature protein. The modified pCGN1761 with the destroyed *SphI* site is cleaved with *PstI* and *EcoRI* and treated with T4 DNA polymerase to render termini blunt. prbcS-8A is cleaved *SphI* and ligated to the annealed molecular adaptor of the sequence described above. The resultant product is 5'-terminally phosphorylated by treatment with T4 kinase. Subsequent cleavage with *BamHI* releases the promoter-transit peptide containing fragment which is treated with T4 DNA polymerase to render the *BamHI* terminus blunt. The promoter-transit peptide fragment thus generated is cloned into the prepared pCGN1761ENX vector, generating a construction comprising the *rbcS-8A* promoter and transit peptide with an *SphI* site located at the cleavage site for insertion of heterologous genes. Further, downstream of the *SphI* site there are *EcoRI* (re-created), *NotI,* and *XhoI* cloning sites. This construction is designated pCGN1761rbcS/CT.

Similar manipulations can be undertaken to utilize other GS2 chloroplast transit peptide encoding sequences from other sources (monocotyledonous and dicotyledonous) and from other genes. In addition, similar procedures can be followed to achieve targeting to other subcellular compartments such as mitochondria.

Example 38: **Techniques for the Isolation of New Promoters Suitable for the Expression of APS Genes**

New promoters are isolated using standard molecular biological techniques including any of the techniques described below. Once isolated, they are fused to reporter genes such as GUS or LUC and their expression pattern in transgenic plants analyzed (Jefferson *et al.*

EMBO J. 6: 3901-3907 (1987); Ow *et al.* Science 234: 856-859 (1986)). Promoters which show the desired expression pattern are fused to APS genes for expression *in planta*.

Subtractive cDNA Cloning

Subtractive cDNA cloning techniques are useful for the generation of cDNA libraries enriched for a particular population of mRNAs (*e.g.* Hara *et al.* Nucl. Acids Res. 19: 1097-7104 (1991)). Recently, techniques have been described which allow the construction of subtractive libraries from small amounts of tissue (Sharma *et al.* Biotechniques 15: 610-612 (1993)). These techniques are suitable for the enrichment of messages specific for tissues which may be available only in small amounts such as the tissue immediately adjacent to wound or pathogen infection sites.

Differential Screening by Standard Plus/Minus Techniques

λ phage carrying cDNAs derived from different RNA populations (*viz.* root versus whole plant, stem specific versus whole plant, local pathogen infection points versus whole plant, *etc.*) are plated at low density and transferred to two sets of hybridization filters (for a review of differential screening techniques see Calvet, Pediatr. Nephrol. 5: 751-757 (1991). cDNAs derived from the "choice" RNA population are hybridized to the first set and cDNAs from whole plant RNA are hybridized to the second set of filters. Plaques which hybridize to the first probe, but not to the second, are selected for further evaluation. They are picked and their cDNA used to screen Northern blots of "choice" RNA versus RNA from various other tissues and sources. Clones showing the required expression pattern are used to clone gene sequences from a genomic library to enable the isolation of the cognate promoter. Between 500 and 5000 bp of the cloned promoter is then fused to a reporter gene (*e.g.* GUS, LUC) and reintroduced into transgenic plants for expression analysis.

Differential Screening by Differential Display

RNA is isolated from different sources *i.e.* the choice source and whole plants as control, and subjected to the differential display technique of Liang and Pardee (Science 257: 967-971 (1992)). Amplified fragments which appear in the choice RNA, but not the control are gel purified and used as probes on Northern blots carrying different RNA samples as described above. Fragments which hybridize selectively to the required RNA are cloned and used as probes to isolate the cDNA and also a genomic DNA fragment from which the

promoter can be isolated. The isolated promoter is fused to a GUS or LUC reporter gene as described above to assess its expression pattern in transgenic plants.

Promoter Isolation Using "Promoter Trap" Technology

The insertion of promoterless reporter genes into transgenic plants can be used to identify sequences in a host plant which drive expression in desired cell types or with a desired strength. Variations of this technique is described by Ott & Chua (Mol. Gen. Genet. 223: 169-179 (1990)) and Kertbundit *et al.* (Proc. Natl. Acad. Sci. USA 88: 5212-5216 (1991)). In standard transgenic experiments the same principle can be extended to identify enhancer elements in the host genome where a particular transgene may be expressed at particularly high levels.

Example 39: **Transformation of Dicotyledons**

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques which do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are described by Paszkowski *et al.*, EMBO J 3: 2717-2722 (1984), Potrykus *et al.*, Mol. Gen. Genet. 199: 169-177 (1985), Reich *et al.*, Biotechnology 4: 1001-1004 (1986), and Klein *et al.*, Nature 327: 70-73 (1987). In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

Agrobacterium-mediated transformation is a preferred technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. The many crop species which are routinely transformable by *Agrobacterium* include tobacco, tomato, sunflower, cotton, oilseed rape, potato, soybean, alfalfa and poplar (EP 0 317 511 (cotton), EP 0 249 432 (tomato, to Calgene), WO 87/07299 (*Brassica*, to Calgene), US 4,795,855 (poplar)). *Agrobacterium* transformation typically involves the transfer of the binary vector carrying the foreign DNA of interest (*e.g.* pCIB200 or pCIB2001) to an appropriate *Agrobacterium* strain which may depend of the complement of *vir* genes carried by the host *Agrobacterium* strain either on a co-resident Ti plasmid or chromosomally (*e.g.* strain CIB542 for pCIB200 and pCIB2001 (Uknes *et al.*

Plant Cell 5: 159-169 (1993)). The transfer of the recombinant binary vector to *Agrobacterium* is accomplished by a triparental mating procedure using *E. coli* carrying the recombinant binary vector, a helper *E. coli* strain which carries a plasmid such as pRK2013 and which is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by DNA transformation (Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877(1988)).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

Example 40: **Transformation of Monocotyledons**

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* co-transformation) and both these techniques are suitable for use with this invention. Co-transformation may have the advantage of avoiding complex vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded desirable. However, a disadvantage of the use of co-transformation is the less than 100% frequency with which separate DNA species are integrated into the genome (Schocher *et al.* Biotechnology 4: 1093-1096 (1986)).

Patent Applications EP 0 292 435 (to Ciba-Geigy), EP 0 392 225 (to Ciba-Geigy) and WO 93/07278 (to Ciba-Geigy) describe techniques for the preparation of callus and protoplasts from an élite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm *et al.* (Plant Cell 2: 603-618 (1990)) and Fromm *et al.* (Biotechnology 8: 833-839 (1990)) have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, application WO 93/07278 (to Ciba-Geigy) and Koziel

et al. (Biotechnology 11: 194-200 (1993)) describe techniques for the transformation of élite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistics device for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has been described for *Japonica*-types and *Indica*-types (Zhang *et al.*, Plant Cell Rep 7: 379-384 (1988); Shimamoto *et al.* Nature 338: 274-277 (1989); Datta *et al.* Biotechnology 8: 736-740 (1990)). Both types are also routinely transformable using particle bombardment (Christou *et al.* Biotechnology 9: 957-962 (1991)).

Patent Application EP 0 332 581 (to Ciba-Geigy) describes techniques for the generation, transformation and regeneration of Pooideae protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation was been described by Vasil *et al.* (Biotechnology 10: 667-674 (1992)) using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al.* (Biotechnology 11: 1553-1558 (1993)) and Weeks *et al.* (Plant Physiol. 102: 1077-1084 (1993)) using particle bombardment of immature embryos and immature embryo-derived callus. A preferred technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, any number of embryos (0.75-1 mm in length) are plated onto MS medium with 3% sucrose (Murashiga & Skoog, Physiologia Plantarum 15: 473-497 (1962)) and 3 mg/l 2,4-D for induction of somatic embryos which is allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and placed onto the osmoticum (*i.e.* induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 h and are then bombarded. Twenty embryos per target plate is typical, although not critical. An appropriate gene-carrying plasmid (such as pCIB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont Biolistics® helium device using a burst pressure of ~1000 psi using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 h (still on

osmoticum). After 24 hrs, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l basta in the case of pCIB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contained half-strength MS, 2% sucrose, and the same concentration of selection agent. Patent application WO 94/13822 describes methods for wheat transformation and is hereby incorporated by reference.

Example 41: **Expression of Pyrrolnitrin in Transgenic Plants**

The GC content of all four pyrrolnitrin ORFs is between 62 and 68% and consequently no AT-content related problems are anticipated with their expression in plants. It may, however, be advantageous to modify the genes to include codons preferred in the appropriate target plant species. Fusions of the kind described below can be made to any desired promoter with or without modification (*e.g.* for optimized translational initiation in plants or for enhanced expression).

Expression behind the 35S Promoter

Each of the four pyrrolnitrin ORFs is transferred to pBluescript KS II for further manipulation. This is done by PCR amplification using primers homologous to each end of each gene and which additionally include a restriction site to facilitate the transfer of the amplified fragments to the pBluescript vector. For ORF1, the aminoterminal primer includes a *SalI* site and the carboxyterminal primer a *NotI* site. Similarly for ORF2, the aminoterminal primer includes a *SalI* site and the carboxyterminal primer a *NotI* site. For ORF3, the aminoterminal primer includes a *NotI* site and the carboxyterminal primer an *XhoI* site. Similarly for ORF4, the aminoterminal primer includes a *NotI* site and the carboxyterminal primer an *XhoI* site. Thus, the amplified fragments are cleaved with the appropriate restriction enzymes (chosen because they do not cleave within the ORF) and are then ligated into pBluescript, also correspondingly cleaved. The cloning of the individual ORFs in pBluescript facilitates their subsequent manipulation.

Destruction of internal restriction sites which are required for further construction is undertaken using the procedure of "inside-outside PCR" (Innes *et al.* PCR Protocols: A guide to methods and applications. Academic Press, New York (1990)). Unique restriction sites sought at either side of the site to be destroyed (ideally between 100 and 500 bp from the site to be destroyed) and two separate amplifications are set up. One extends from the unique site left of the site to be destroyed and amplifies DNA up to the site to be destroyed with an amplifying oligonucleotide which spans this site and incorporates an appropriate base change. The second amplification extends from the site to be destroyed up to the unique site rightwards of the site to be destroyed. The oligonucleotide spanning the site to be destroyed in this second reaction incorporates the same base change as in the first amplification and ideally shares an overlap of between 10 and 25 nucleotides with the oligonucleotide from the first reaction. Thus the products of both reactions share an overlap which incorporates the same base change in the restriction site corresponding to that made in each amplification. Following the two amplifications, the amplified products are gel purified (to remove the four oligonucleotide primers used), mixed together and reamplified in a PCR reaction using the two primers spanning the unique restriction sites. In this final PCR reaction the overlap between the two amplified fragments provides the priming necessary for the first round of synthesis. The product of this reactions extends from the leftwards unique restriction site to the rightwards unique restriction site and includes the modified restriction site located internally. This product can be cleaved with the unique sites and inserted into the unmodified gene at the appropriate location by replacing the wild-type fragment.

To render ORF1 free of the first of its two internal *SphI* sites oligonucleotides spanning and homologous to the unique *XmaI* and *EspI* are designed. The *XmaI* oligonucleotide is used in a PCR reaction together with an oligonucleotide spanning the first *SphI* site and which comprises the sequence ....CCCCC<u>T</u>CATGC.... (lower strand, SEQ ID NO:15), thus introducing a base change into to *SphI* site. A second PCR reaction utilizes an oligonucleotide spanning the *SphI* site (upper strand) comprising the sequence ....GCATG<u>A</u>GGGGG.... (SEQ ID NO:16) and is used in combination with the *EspI* site-spanning oligonucleotide. The two products are gel purified and themselves amplified with the *XmaI* and *EspI*-spanning oligonucleotides and the resultant fragment is cleaved with *XmaI* and *EspI* and used to replace the native fragment in the ORF1 clone. According to

the above description, the modified *SphI* site is GCATGA and does not cause a codon change. Other changes in this site are possible (*i.e.* changing the second nucleotide to a G, T, or A) without corrupting amino acid integrity.
A similar strategy is used to destroy the second *SphI* site in ORF1. In this case, *EspI* is a suitable leftwards-located restriction site, and the rightwards-located restriction site is *PstI*, located close to the 3' end of the gene or alternatively *SstI* which is not found in the ORF sequence, but immediately adjacent in the pBluescript polylinker. In this case an appropriate oligonucleotide is one which spans this site, or alternatively one of the available pBluescript sequencing primers. This *SphI* site is modified to GAATGC or GCATGT or GAATGT. Each of these changes destroys the site without causing a codon change.

To render ORF2 free of its single *SphI* site a similar procedure is used. Leftward restriction sites are provided by *PstI* or *MluI*, and a suitable rightwards restriction site is provided by *SstI* in the pBluescript polylinker. In this case the site is changed to GCTTGC, GCATGC or GCTTGT; these changes maintain amino acid integrity.

ORF3 has no internal *SphI* sites.

In the case of ORF4, *PstI* provides a suitable rightwards unique site, but there is no suitable site located leftwards of the single *SphI* site to be changed. In this case a restriction site in the pBluescript polylinker can be used to the same effect as already described above. The *SphI* site is modified to GGATGC, GTATGC, GAATGC, or GCATGT *etc.*.

The removal of *SphI* sites from the pyrrolnitrin biosynthetic genes as described above facilitates their transfer to the pCGN1761SENX vector by amplification using an aminoterminal oligonucleotide primer which incorporates an *SphI* site at the ATG and a carboxyterminal primer which incorporates a restriction site not found in the gene being amplified. The resultant amplified fragment is cleaved with *SphI* and the restriction enzyme cutting the carboxyterminal sequence and cloned into pCGN1761SENX. Suitable restriction enzyme sites for incorporation into the carboxyterminal primer are *NotI* (for all four ORFs), *XhoI* (for ORF3 and ORF4), and *EcoRI* (for ORF4). Given the requirement for the nucleotide C at position 6 within the *SphI* recognition site, in some cases the second codon of the ORF may require changing so as to start with the nucleotide C. This construction

fuses each ORF at its ATG to the *SphI* sites of the translation-optimized vector pCGN1761SENX in operable linkage to the double 35S promoter. After construction is complete the final gene insertions and fusion points are resequenced to ensure that no undesired base changes have occurred.

By utilizing an aminoterminal oligonucleotide primer which incorporates an *NcoI* site at its ATG instead of an *SphI* site, ORFs 1-4 can also be easily cloned into to the translation-optimized vector pCGN1761NENX. None of the four pyrrolnitrin biosynthetic gene ORFs carry an *NcoI* site and consequently there is no requirement in this case to destroy internal restriction sites. Primers for the carboxyterminus of the gene are designed as described above and the cloning is undertaken in a similar fashion. Given the requirement for the nucleotide G at position 6 within the *NcoI* recognition site, in some cases the second codon of the ORF may require changing so as to start with the nucleotide G. This construction fuses each ORF at its ATG to the *NcoI* site of pCGN1761NENX in operable linkage to the double 35S promoter.

The expression cassettes of the appropriate pCGN1761-derivative vectors are transferred to transformation vectors. Where possible multiple expression cassettes are transferred to a single transformation vector so as to reduce the number of plant transformations and crosses between transformants which may be required to produce plants expressing all four ORFs and thus producing pyrrolnitrin.

Expression behind 35S with Chloroplast Targeting

The pyrrolnitrin ORFs 1-4 amplified using oligonucleotides carrying an *SphI* site at their aminoterminus are cloned into the 35S-chloroplast targeted vector pCGN1761/CT. The fusions are made to the *SphI* site located at the cleavage site of the *rbcS* transit peptide. The expression cassettes thus created are transferred to appropriate transformation vectors (see above) and used to generate transgenic plants. As tryptophan, the precursor for pyrrolnitrin biosynthesis, is synthesized in the chloroplast, it may be advantageous to express the biosynthetic genes for pyrrolnitrin in the chloroplast to ensure a ready supply of substrate. Transgenic plants expressing all four ORFs will target all four gene products to the chloroplast and will thus synthesize pyrrolnitrin in the chloroplast.

Expression behind *rbcS* with Chloroplast Targeting

The pyrrolnitrin ORFs 1-4 amplified using oligonucleotides carrying an *SphI* site at their aminoterminus are cloned into the *rbcS*-chloroplast targeted vector pCGN1761rbcS/CT. The fusions are made to the *SphI* site located at the cleavage site of the *rbcS* transit peptide. The expression cassettes thus created are transferred to appropriate transformation vectors (see above) and used to generate transgenic plants. As tryptophan, the precursor for pyrrolnitrin biosynthesis, is synthesized in the chloroplast, it may be advantageous to express the biosynthetic genes for pyrrolnitrin in the chloroplast to ensure a ready supply of substrate. Transgenic plants expressing all four ORFs will target all four gene products to the chloroplast and will thus synthesize pyrrolnitrin in the chloroplast. The expression of the four ORFs will, however, be light induced.

Example 42: **Expression of Soraphen in Transgenic Plants**

Clone p98/1 contains the entirety of the soraphen biosynthetic gene ORF1 which encodes five biosynthetic modules for soraphen biosynthesis. The partially sequenced ORF2 contains the remaining three modules, and further required for soraphen biosynthesis is the soraphen methylase located on the same operon.

Soraphen ORF1 is manipulated for expression in transgenic plants in the following manner. A DNA fragment is amplified from the aminoterminus of ORF1 using PCR and p98/1 as template. The 5' oligonucleotide primer includes either an *SphI* site or an *NcoI* site at the ATG for cloning into the vectors pCGN1761SENX or pCGNNENX respectively. Further, the 5' oligonucleotide includes either the base C (for *SphI* cloning) or the base G (for *NcoI* cloning) immediately after the ATG, and thus the second amino acid of the protein is changed either to a histidine or an aspartate (other amino acids can be selected for position 2 by additionally changing other bases of the second codon). The 3' oligonucleotide for the amplification is located at the first *BglII* site of the ORF and incorporates a distal *EcoRI* site enabling the amplified fragment to be cleaved with *SphI* (or *NcoI*) and *EcoRI*, and then cloned into pCGN1761SENX (or pCGN1761NENX). To facilitate cleavage of the amplified fragments, each oligonucleotide includes several additional bases at its 5' end. The oligonucleotides preferably have 12-30 bp homology to the ORF1 template, in addition to the required restriction sites and additional sequences. This manipulation fuses the aminoterminal ~112 amino acids of ORF1 at its ATG to the *SphI* or *NcoI* sites of the

translation optimized vectors pCGN1761SENX or pCGN1761NENX in linkage to the double 35S promoter. The remainder of ORF1 is carried on three *BglII* fragments which can be sequentially cloned into the unique *BglII* site of the above-detailed constructions. The introduction of the first of these fragments is no problem, and requires only the cleavage of the aminoterminal construction with *BglII* followed by introduction of the first of these fragments. For the introduction of the two remaining fragments, partial digestion of the aminoterminal construction is required (since this construction now has an additional *BglII* site), followed by introduction of the next *BglII* fragment. Thus, it is possible to construct a vector containing the entire ~25 kb of soraphen ORF1 in operable fusion to the 35S promoter.

An alternative approach to constructing the soraphen ORF1 by the fusion of sequential restriction fragments is to amplify the entire ORF using PCR. Barnes (Proc. Natl. Acad. Sci USA 91: 2216-2220 (1994)) has recently described techniques for the high-fidelity amplification of fragments by PCR of up to 35 kb, and these techniques can be applied to ORF1. Oligonucleotides specific for each end of ORF1, with appropriate restriction sites added are used to amplify the entire coding region, which is then cloned into appropriate sites in a suitable vector such as pCGN1761 or its derivatives. Typically after PCR amplification, resequencing is advised to ensure that no base changes have arisen in the amplified sequence. Alternatively, a functional assay can be done directly in transgenic plants.

Yet another approach to the expression of the genes for polyketide biosynthesis (such as soraphen) in transgenic plants is the construction, for expression in plants, of transcriptional units which comprise less than the usual complement of modules, and to provide the remaining modules on other transcriptional units. As it is believed that the biosynthesis of polyketide antibiotics such as soraphen is a process which requires the sequential activity of specific modules and that for the synthesis of a specific molecule these activities should be provided in a specific sequence, it is likely that the expression of different transgenes in a plant carrying different modules may lead to the biosynthesis of novel polyketide molecules because the sequential enzymatic nature of the wild-type genes is determined by their configuration on a single molecule. It is assumed that the localization of five specific modules for soraphen biosynthesis on ORF1 is determinatory in the biosynthesis of

soraphen, and that the expression of, say three modules on one transgene and the other two on another, together with ORF2, may result in biosynthesis of a polyketide with a different molecular structure and possibly with a different antipathogenic activity. This invention encompasses all such deviations of module expression which may result in the synthesis in transgenic organisms of novel polyketides.

Although specific construction details are only provided for ORF1 above, similar techniques are used to express ORF2 and the soraphen methylase in transgenic plants. For the expression of functional soraphen in plants it is anticipated that all three genes must be expressed and this is done as detailed in this specification.

Fusions of the kind described above can be made to any desired promoter with or without modification (*e.g.* for optimized translational initiation in plants or for enhanced expression). As the ORFs identified for soraphen biosynthesis are around 70% GC rich it is not anticipated that the coding sequences should require modification to increase GC content for optimal expression in plants. It may, however, be advantageous to modify the genes to include codons preferred in the appropriate target plant species.

Example 43: **Expression of Phenazine in Transgenic Plants**

The GC content of all the cloned genes encoding biosynthetic enzymes for phenazine synthesis is between 58 and 65% and consequently no AT-content related problems are anticipated with their expression in plants (although it may be advantageous to modify the genes to include codons preferred in the appropriate target plant species.). Fusions of the kind described below can be made to any desired promoter with or without modification (*e.g.* for optimized translational initiation in plants or for enhanced expression).

Expression behind the 35S Promoter

Each of the three phenazine ORFs is transferred to pBluescript SK II for further manipulation. The *phzB* ORF is transferred as an *EcoRI-BglII* fragment cloned from plasmid pLSP18-6H3del3 containing the entire phenazine operon. This fragment is transferred to the *EcoRI-BamHI* sites of pBluescript SK II. The *phzC* ORF is transferred from pLSP18-6H3del3 as an *XhoI-ScaI* fragment cloned into the *XhoI-SmaI* sites of

pBluescript II SK. The *phzD* ORF is transferred from pLSP18-6H3del3 as a *BglII-HindIII* fragment into the *BamHI-HindIII* sites of pBluescript II SK.
Destruction of internal restriction sites which are required for further construction is undertaken using the procedure of "inside-outside PCR" described above (Innes *et al.* PCR Protocols: A guide to methods and applications. Academic Press, New York (1990)). In the case of the *phzB* ORF two *SphI* sites are destroyed (one site located upstream of the ORF is left intact). The first of these is destroyed using the unique restriction sites *EcoRI* (left of the *SphI* site to be destroyed) and *BclI* (right of the *SphI* site). For this manipulation to be successful, the DNA to be *BclI* cleaved for the final assembly of the inside-outside PCR product must be produced in a *dam-minus E. coli* host such as SCS110 (Stratagene). For the second *phzB SphI* sites, the selected unique restriction sites are *PstI* and *SpeI*, the latter being beyond the *phzB* ORF in the pBluescript polylinker. The *phzC* ORF has no internal *SphI* sites, and so this procedure is not required for *phzC*. The *phzD* ORF, however, has a single *SphI* site which can be removed using the unique restriction sites *XmaI* and *HindIII* (the *XmaI/SmaI* site of the pBluescript polylinker is no longer present due to the insertion of the ORF between the *BamHI* and *HindIII* sites).

The removal of *SphI* sites from the phenazine biosynthetic genes as described above facilitates their transfer to the pCGN1761SENX vector by amplification using an aminoterminal oligonucleotide primer which incorporates an *SphI* site at the ATG and a carboxyterminal primer which incorporates a restriction site not found in the gene being amplified. The resultant amplified fragment is cleaved with *SphI* the restriction enzyme cutting the carboxyterminal sequence and cloned into pCGN1761SENX. Suitable restriction enzyme sites for incorporation into the carboxyterminal primer are *EcoRI* and *NotI* (for all three ORFs; *NotI* will need checking when sequence complete), and XhoI (for phzB and phzD). Given the requirement for the nucleotide C at position 6 within the SphI recognition site, in some cases the second codon of the ORF may require changing so as to start with the nucleotide C. This construction fuses each ORF at its ATG to the SphI sites of the translation-optimized vector pCGN1761SENX in operable linkage to the double 35S promoter. After construction is complete the final gene insertions and fusion points are resequenced to ensure that no undesired base changes have occurred.

By utilizing an aminoterminal oligonucleotide primer which incorporates an *NcoI* site at its ATG instead of an *SphI* site, the three *phz* ORFs can also be easily cloned into to the translation-optimized vector pCGN1761NENX. None of the three phenazine biosynthetic gene ORFs carry an *NcoI* site and consequently there is no requirement in this case to destroy internal restriction sites. Primers for the carboxyterminus of the gene are designed as described above and the cloning is undertaken in a similar fashion. Given the requirement for the nucleotide G at position 6 within the NcoI recognition site, in some cases the second codon of the ORF may require changing so as to start with the nucleotide G. This construction fuses each ORF at its ATG to the *NcoI* site of pCGN1761NENX in operable linkage to the double 35S promoter.

The expression cassettes of the appropriate pCGN1761-derivative vectors are transferred to transformation vectors. Where possible multiple expression cassettes are transferred to a single transformation vector so as to reduce the number of plant transformations and crosses between transformants which may be required to produce plants expressing all four ORFs and thus producing phenazine.

Expression behind 35S with Chloroplast Targeting

The three phenazine ORFs amplified using oligonucleotides carrying an *SphI* site at their aminoterminus are cloned into the 35S-chloroplast targeted vector pCGN1761/CT. The fusions are made to the *SphI* site located at the cleavage site of the *rbcS* transit peptide. The expression cassettes thus created are transferred to appropriate transformation vectors (see above) and used to generate transgenic plants. As chorismate, the likely precursor for phenazine biosynthesis, is synthesized in the chloroplast, it may be advantageous to express the biosynthetic genes for phenazine in the chloroplast to ensure a ready supply of substrate. Transgenic plants expressing all three ORFs will target all three gene products to the chloroplast and will thus synthesize phenazine in the chloroplast.

Expression behind *rbcS* with Chloroplast Targeting

The three phenazine ORFs amplified using oligonucleotides carrying an *SphI* site at their aminoterminus are cloned into the *rbcS*-chloroplast targeted vector pCGN1761rbcS/CT. The fusions are made to the *SphI* site located at the cleavage site of the *rbcS* transit peptide. The expression cassettes thus created are transferred to appropriate

transformation vectors (see above) and used to generate transgenic plants. As chorismate, the likely precursor for phenazine biosynthesis, is synthesized in the chloroplast, it may be advantageous to express the biosynthetic genes for phenazine in the chloroplast to ensure a ready supply of substrate. Transgenic plants expressing all three ORFs will target all four gene products to the chloroplast and will thus synthesize phenazine in the chloroplast. The expression of the three ORFs will, however, be light induced.

Example 44: **Expression of the Non-Ribosomally Synthesized Peptide Antibiotic Gramicidin in Transgenic Plants**

The three *Bacillus brevis* gramicidin biosynthetic genes *grsA, grsB* and *grsT* have been previously cloned and sequenced (Turgay *et al.* Mol. Microbiol. 6: 529-546 (1992); Kraetzschmar *et al.* J. Bacteriol. 171: 5422-5429 (1989)). They are 3296, 13358, and 770 bp in length, respectively. These sequences are also published as GenBank accession numbers X61658 and M29703. The manipulations described here can be undertaken using the publicly available clones published by Turgay *et al.* (*supra*) and Kraetzschmar *et al.* (*supra*), or alternatively from newly isolated clones from *Bacillus brevis* isolated as described herein.

Each of the three ORFs *grsA, grsB,* and *grsT* is PCR amplified using oligonucleotides which span the entire coding sequence. The leftward (upstream) oligonucleotide includes an *SstI* site and the rightward (downstream) oligonucleotide includes an *XhoI* site. These restriction sites are not found within any of the three coding sequences and enable the amplified products to be cleaved with *SstI* and *XhoI* for insertion into the corresponding sites of pBluescript II SK. This generates the clones pBL-GRSa, pBLGRSb and pBLGRSt. The CG content of these genes lies between 35 and 38%. Ideally, the coding sequences encoding the three genes may be remade using the techniques referred to in Section K, however it is possible that the unmodified genes may be expressed at high levels in transgenic plants without encountering problems due to their AT content. In any case it may be advantageous to modify the genes to include codons preferred in the appropriate target plant species.

The ORF *grsA* contains no *SphI* site and no *NcoI* site. This gene can be thus amplified from pBLGSRa using an aminoterminal oligonucleotide which incorporates either an *SphI*

site or an *NcoI* site at the ATG, and a second carboxyterminal oligonucleotide which incorporates an *XhoI* site, thus enabling the amplification product to be cloned directly into pCGN1761SENX or pCGN1761NENX behind the double 35S promoter.
The ORF *grsB* contains no *NcoI* site and therefore this gene can be amplified using an aminoterminal oligonucleotide containing an *NcoI* site in the same way as described above for the *grsA* ORF; the amplified fragment is cleaved with *NcoI* and *XhoI* and ligated into pCGN1761NENX. However, the *grsB* ORF contains three *SphI* sites and these are destroyed to facilitate the subsequent cloning steps. The sites are destroyed using the "inside-outside" PCR technique described above. Unique cloning sites found within the *grsB* gene but not within pBluescript II SK are *EcoN1, PflM1*, and *RsrII*. Either *EcoN1* or *PflM1* can be used together with *RsrII* to remove the first two sites and *RsrII* can be used together with the *ApaI* site of the pBluescript polylinker to remove the third site. Once these sites have been destroyed (without causing a change in amino acid), the entirety of the *grsB* ORF can be amplified using an aminoterminal oligonucleotide including an *SphI* site at the ATG and a carboxyterminal oligonucleotide incorporating an *XhoI* site. The resultant fragment is cloned into pCGN1761SENX. In order to successfully PCR-amplify fragments of such size, amplification protocols are modified in view of Barnes (1994, Proc. Natl. Acad. Sci USA 91: 2216-2220 (1994)) who describes the high fidelity amplification of large DNA fragments. An alternative approach to the transfer of the *grsB* ORF to pCGN1761SENX without necessitating the destruction of the three *SphI* restriction sites involves the transfer to the *SphI* and *XhoI* cloning sites of pCGN1761SENX of an aminoterminal fragment of *grsB* by amplification from the ATG of the gene using an aminoterminal oligonucleotide which incorporates a *SphI* site at the ATG, and a second oligonucleotide which is adjacent and 3' to the *PflM1* site in the ORF and which includes an *XhoI* site. Thus the aminoterminal amplified fragment is cleaved with *SphI* and *XhoI* and cloned into pCGN1761SENX. Subsequently the remaining portion of the *grsB* gene is excised from pBLGRSb using *PflMI* and *XhoI* (which cuts in the pBluescript polylinker) and cloned into the aminoterminal carrying construction cleaved with *PflMI* and *XhoI* to reconstitute the gene.

The ORF *grsT* contains no *SphI* site and no *NcoI* site. This gene can be thus amplified from pBLGSRt using an aminoterminal oligonucleotide which incorporates either an *SphI* site or an *NcoI* site at the initiating codon which is changed to ATG (from GTG) for

expression in plants, and a second carboxyterminal oligonucleotide which incorporates an *XhoI* site, thus enabling the amplification product to be cloned directly into pCGN1761SENX or pCGN1761NENX behind the double 35S promoter.
Given the requirement for the nucleotide C at position 6 within the *SphI* recognition site, and the requirement for the nucleotide G at position 6 within the *NcoI* recognition site, in some cases the second codon of the ORF may require changing so as to start with the appropriate nucleotide.

Transgenic plants are created which express all three gramicidin biosynthetic genes as described elsewhere in the specification. Transgenic plants expressing all three genes synthesize gramicidin.

Example 45: **Expression of the Ribosomally Synthesized Peptide Lantibiotic Epidermin in Transgenic Plants**

The *epiA* ORF encodes the structural unit for epidermin biosynthesis and is approximately 420 bp in length (GenBank Accession No. X07840; Schnell *et al.* Nature 333: 276-278 (1988)). This gene can be subcloned using PCR techniques from the plasmid pTü32 into pBluescript SK II using oligonucleotides carrying the terminal restriction sites *BamHI* (5') and *PstI* (3'). The *epiA* gene sequence has a GC content of 27% and this can be increased using techniques of gene synthesis referred to elsewhere in this specification; this sequence modification may not be essential, however, to ensure high-level expression in plants. Subsequently the *epiA* ORF is transferred to the cloning vector pCGN1761SENX or pCGN1761NENX by PCR amplification of the gene using an aminoterminal oligonucelotide spanning the initiating methionine and carrying an *SphI* site (for cloning into pCGN1761SENX) or an *NcoI* site (for cloning into pCGN1761NENX), together with a carboxyterminal oligonucleotide carrying an *EcoRI*, a *NotI*, or an *XhoI* site for cloning into either pCGN1761SENX or pCGN1761NENX. Given the requirement for the nucleotide C at position 6 within the *SphI* recognition site, and the requirement for the nucleotide G at position 6 within the *NcoI* recognition site, in some cases the second codon of the ORF may require changing so as to start with the appropriate nucleotide.

Using cloning techniques described in this specification or well known in the art, the remaining genes of the *epi* operon (*viz. epiB, epiC, epiD, epiQ,* and *epiP*) are subcloned

from plasmid pTü32 into pBluescript SK II. These genes are responsible for the modification and polymerization of the *epiA*-encoded structural unit and are described in Kupke *et al.* (J. Bacteriol. 174: 5354-5361 (1992)) and Schnell *et al.* (Eur. J. Biochem. 204: 57-68 (1992)). The subcloned ORFs are manipulated for transfer to pCGN1761-derivative vectors as described above. The expression cassettes of the appropriate pCGN1761-derivative vectors are transferred to transformation vectors. Where possible multiple expression cassettes are transferred to a single transformation vector so as to reduce the number of plant transformations and crosses between transformants which may be required to produce plants expressing all required ORFs and thus producing epidermin.

L. Analysis of Transgenic Plants for APS Accumulation

Example 46: **Analysis of APS Gene Expression**

Expression of APS genes in transgenic plants can be analyzed using standard Northern blot techniques to assess the amount of APS mRNA accumulating in tissues. Alternatively, the quantity of APS gene product can be assessed by Western analysis using antisera raised to APS biosynthetic gene products. Antisera can be raised using conventional techniques and proteins derived from the expression of APS genes in a host such as *E. coli*. To avoid the raising of antisera to multiple gene products from *E. coli* expressing multiple APS genes from multiple ORF operons, the APS biosynthetic genes can be expressed individually in *E. coli*. Alternatively, antisera can be raised to synthetic peptides designed to be homologous or identical to known APS biosynthetic predicted amino acid sequence. These techniques are well known in the art.

Example 47: **Analysis of APS Production in Transgenic Plants**

For each APS, known protocols are used to detect production of the APS in transgenic plant tissue. These protocols are available in the appropriate APS literature. For pyrrolnitrin, the procedure described in example 11 is used, and for soraphen the procedure described in example 17. For phenazine determination, the procedure described in example 18 can be used. For non-ribosomal peptide antibiotics such as gramicidin S, an appropriate general technique is the assaying of ATP-PP_i exchange. In the case of gramicidin, the *grsA* gene can be assayed by phenylalanine-dependent ATP-PP_i exchange

and the *grsB* gene can be assayed by proline, valine, ornithine, or leucine-dependent ATP-PP_i exchange. Alternative techniques are described by Gause & Brazhnikova (Lancet 247: 715 (1944)). For ribosomally synthesized peptide antibiotics isolation can be achieved by butanol extraction, dissolving in methanol and diethyl ether, followed by chromatography as described by Allgaier *et al.* for epidermin (Eur. Ju. Biochem. 160: 9-22 (1986)). For many APSs (*e.g.* pyrrolnitrin, gramicidin, phenazine) appropriate techniques are provided in the Merck Index (Merck & Co., Rahway, NJ (1989)).

M. Assay of Disease Resistance in Transgenic Plants

Transgenic plants expressing APS biosynthetic genes are assayed for resistance to phytopathogens using techniques well known in phytopathology. For foliar pathogens, plants are grown in the greenhouse and at an appropriate stage of development inoculum of a phytopathogen of interest is introduced at in an appropriate manner. For soil-borne phytopathogens, the pathogen is normally introduced into the soil before or at the time the seeds are planted. The choice of plant cultivar selected for introduction of the genes will have taken into account relative phytopathogen sensitivity. Thus, it is preferred that the cultivar chosen will be susceptible to most phytopathogens of interest to allow a determination of enhanced resistance.

Assay of Resistance to Foliar Phytopathogens

Example 48: **Disease Resistance to Tobacco Foliar Phytopathogens**

Transgenic tobacco plants expressing APS genes and shown to poduce APS compound are subjected to the following disease tests.

***Phytophthora parasitica*/Black shank** Assays for resistance to *Phytophthora parasitica,* the causative organism of black shank are performed on six-week-old plants grown as described in Alexander *et al.*, Pro. Natl. Acad. Sci. USA 90: 7327-7331. Plants are watered, allowed to drain well, and then inoculated by applying 10 mL of a sporangium suspension (300 sporangia/mL) to the soil. Inoculated plants are kept in a greenhouse maintained at 23-25 C day temperature, and 20-22 C night temperature. The wilt index used for the assay is as follows: 0 = no symptoms; 1 = some sign of wilting, with reduced turgidity; 2 = clear wilting symptoms, but no rotting or stunting; 3 = clear wilting symptoms with stunting, but no apparent stem rot; 4 = severe wilting, with visible stem rot and some damage to root

system; 5 = as for 4, but plants near death or dead, and with severe reduction of root system. All assays are scored blind on plants arrayed in a random design.

Pseudomonas syringae *Pseudomonas syringae* pv. tabaci (strain #551) is injected into the two lower leaves of several 6-7 week old plants at a concentration of 10^6 or 3×10^6 per ml in H_2O. Six individual plants are evaluated at each time point. *Pseudomonas tabaci* infected plants are rated on a 5 point disease severity scale, 5 = 100% dead tissue, 0 = no symptoms. A T-test (LSD) is conducted on the evaluations for each day and the groupings are indicated after the Mean disease rating value. Values followed by the same letter on that day of evaluation are not statistically significantly different.

Cercospora nicotianae A spore suspension of *Cercospora nicotianae* (ATCC #18366) (100,000-150,000 spores per ml) is sprayed to imminent run-off on to the surface of the leaves. The plants are maintained in 100% humidity for five days. Thereafter the plants are misted with H_2O 5-10 times per day. Six individual plants are evaluated at each time point. *Cercospora nicotianae* is rated on a % leaf area showing disease symptoms basis. A T-test (LSD) is conducted on the evaluations for each day and the groupings are indicated after the Mean disease rating value. Values followed by the same letter on that day of evaluation are not statistically significantly different.

Statistical Analyses All tests include non-transgenic plants (six plants per assay, or the same cultivar as the transgenic lines) (Alexander *et al.*, Pro. Natl. Acad. Sci. USA 90: 7327-7331). Pairwise T-tests are performed to compare different genotype and treatment groups for each rating date.

Assay of Resistance to Soil-Borne Phytopathogens

Example 49: **Resistance to Rhizoctonia solani**

Plant assays to determine resistance to *Rhizoctonia solani* are conducted by planting or transplanting seeds or seedlings into naturally or artificially infested soil. To create artificially infested soil, millet, rice, oat, or other similar seeds are first moistened with water, then autoclaved and inoculated with plugs of the fungal phytopathogen taken from an agar plate. When the seeds are fully overgrown with the phytopathogen, they are air-dried and

ground into a powder. The powder is mixed into soil at a rate experimentally determined to cause disease. Disease may be assessed by comparing stand counts, root lesions ratings, and shoot and root weights of transgenic and non-transgenic plants grown in the infested soil. The disease ratings may also be compared to the ratings of plants grown under the same conditions but without phytopathogen added to the soil.

Example 50: **Resistance to *Pseudomonas solanacearum***

Plant assays to determine resistance to *Pseudomonas solanacearum* are conducted by planting or transplanting seeds or seedlings into naturally or artificially infested soil. To create artificially infested soil, bacteria are grown in shake flask cultures, then mixed into the soil at a rate experimentally determined to cause disease. The roots of the plants may need to be slightly wounded to ensure disease development. Disease may be assessed by comparing stand counts, degree of wilting and shoot and root weights of transgenic and non-transgenic plants grown in the infested soil. The disease ratings may also be compared to the ratings of plants grown under the same conditions but without phytopathogen added to the soil.

Example 51: **Resistance to Soil-Borne Fungi which are Vectors for Virus Transmission**

Many soil-borne *Polymyxa, Olpidium* and *Spongospora* species are vectors for the transmission of viruses. These include (1) *Polymyxa betae* which transmits Beet Necrotic Yellow Vein Virus (the causative agent of rhizomania disease) to sugar beet, (2) *Polymyxa graminis* which transmits Wheat Soil-Borne Mosaic Virus to wheat, and Barley Yellow Mosaic Virus and Barley Mild Mosaic Virus to barley, (3) *Olpidium brassicae* which transmits Tobacco Necrosis Virus to tobacco, and (4) *Spongospora subterranea* which transmits Potato Mop Top Virus to potato. Seeds or plants expressing APSs in their roots (*e.g.* constitutively or under root specific expression) are sown or transplanted in sterile soil and fungal inocula carrying the virus of interest are introduced to the soil. After a suitable time period the transgenic plants are assayed for viral symptoms and accumulation of virus by ELISA and Northern blot. Control experiments involve no inoculation, and inoculation with fungus which does not carry the virus under investigation. The transgenic plant lines under analysis should ideally be susceptible to the virus in order to test the efficacy of the APS-based protection. In the case of viruses such as Barley Mild Mosaic Virus which are both

Polymyxa-transmitted and mechanically transmissible, a further control is provided by the successful mechanical introduction of the virus into plants which are protected against soil-infection by APS expression in roots.

Resistance to virus-transmitting fungi offered by expression of APSs will thus prevent virus infections of target crops thus improving plant health and yield.

Example 52: **Resistance to Nematodes**

Transgenic plants expressing APSs are analyzed for resistance to nematodes. Seeds or plants expressing APSs in their roots (*e.g.* constitutively or under root specific expression) are sown or transplanted in sterile soil and nematode inocula carrying are introduced to the soil. Nematode damage is assessed at an appropriate time point. Root knot nematodes such as *Meloidogyne* spp. are introduced to transgenic tobacco or tomato expressing APSs. Cyst nematodes such as *Heterodera* spp. are introduced to transgenic cereals, potato and sugar beet. Lesion nematodes such as *Pratylenchus* spp. are introduced to transgenic soybean, alfalfa or corn. Reniform nematodes such as *Rotylenchulus* spp. are introduced to transgenic soybean, cotton, or tomato. *Ditylenchus* spp. are introduced to transgenic alfalfa. Detailed techniques for screening for resistance to nematodes are provided in Starr (Ed.; Methods for Evaluating Plant Species for resistance to Plant Parasitic Nematodes, Society of Nematologists, Hyattsville, Maryland (1990))

<u>Examples of Important Phytopathogens in Agricultural Crop Species</u>

Example 53: **Disease Resistance in Maize**

Transgenic maize plants expressing APS genes and shown to poduce APS compound are subjected to the following disease tests. Tests for each phytopathogen are conducted according to standard phytopathological procedures.

Leaf Diseases and Stalk Rots

(1) Northern Corn Leaf Blight (*Helminthosporium turcicum†* syn. *Exserohilum turcicum*).

(2) Anthracnose (*Colletotrichum graminicola†*-same as for Stalk Rot)

(3) Southern Corn Leaf Blight (*Helminthosporium maydis†* syn. *Bipolaris maydis*).

(4) Eye Spot (*Kabatiella zeae*)

(5) Common Rust (*Puccinia sorghi*).

(6) Southern Rust (*Puccinia polysora*).

(7) Gray Leaf Spot (*Cercospora zeae-maydis† and C. sorghi*)

(8) Stalk Rots (a complex of two or more of the following pathogens-*Pythium aphanidermatum†*-early, *Erwinia chrysanthemi-zeae*-early, *Colletotrichum graminicola†, Diplodia maydis†, D. macrospora, Gibberella zeae†, Fusarium moniliforme†, Macrophomina phaseolina, Cephalosporium acremonium*)

(9) Goss' Disease (*Clavibacter nebraskanense*)

Important-Ear Molds

(1) Gibberella Ear Rot (*Gibberella zeae†*-same as for Stalk Rot) *Aspergillus flavus, A. parasiticus.* Aflatoxin

(2) Diplodia Ear Rot (*Diplodia maydis† and D. macrospora*-same organisms as for Stalk Rot)

(3) Head Smut (*Sphacelotheca reiliana*--syn. *Ustilago reiliana*)

Example 54: **Disease Resistance in Wheat**

Transgenic wheat plants expressing APS genes and shown to poduce APS compound are subjected to the following disease tests. Tests for each pathogen are conducted according to standard phytopathological procedures.

(1) Septoria Diseases (*Septoria tritici, S. nodorum*)

(2) Powdery Mildew (*Erysiphe graminis*)

(3) Yellow Rust (*Puccinia striiformis*)

(4) Brown Rust (*Puccinia recondita, P. hordei*)

(5) Others-Brown Foot Rot/Seedling Blight (*Fusarium culmorum* and *Fusarium roseum*), Eyespot (*Pseudocercosporella herpotrichoides*), Take-All (*Gaeumannomyces graminis*)

(6) Viruses (barley yellow mosaic virus, barley yellow dwarf virus, wheat yellow mosaic virus).

N. <u>Assay of Biocontrol Efficacy in Microbial Strains Expressing APS Genes</u>

Example 55: **Protection of Cotton against *Rhizoctonia solani***

Assays to determine protection of cotton from infection caused by *Rhizoctonia solani* are conducted by planting seeds treated with the biocontrol strain in naturally or artificially

infested soil. To create artificially infested soil, millet, rice, oat, or other similar seeds are first moistened with water, then autoclaved and inoculated with plugs of the fungal pathogen taken from an agar plate. When the seeds are fully overgrown with the pathogen, they are air-dried and ground into a powder. The powder is mixed into soil at a rate experimentally determined to cause disease. This infested soil is put into pots, and seeds are placed in furrows 1.5cm deep. The biocontrol strains are grown in shake flasks in the laboratory. The cells are harvested by centrifugation, resuspended in water , and then drenched over the seeds. Control plants are drenched with water only. Disease may be assessed 14 days later by comparing stand counts and root lesions ratings of treated and nontreated seedlings. The disease ratings may also be compared to the ratings of seedlings grown under the same conditions but without pathogen added to the soil.

Example 56: **Protection of Potato against *Claviceps michiganese* subsp. *speedonicum***

Claviceps michiganese subsp. *speedonicum* is the causal agent of potato ring rot disease and is typically spread before planting when "seed" potato tubers are knife cut to generate more planting material. Transmission of the pathogen on the surface of the knife results in the inoculation of entire "seed" batches. Assays to determine protection of potato from the causal agent of ring rot disease are conducted by inoculating potato seed pieces with both the pathogen and the biocontrol strain. The pathogen is introduced by first cutting a naturally infected tuber, then using the knife to cut other tubers into seed pieces. Next, the seed pieces are treated with a suspension of biocontrol bacteria or water as a control. Disease is assessed at the end of the growing season by evaluating plant vigor, yield, and number of tubers infected with *Clavibacter*.

O. Isolation of APSs from Organisms Expressing the Cloned Genes

Example 57: **Extraction Procedures for APS Isolation**

Active APSs can be isolated from the cells or growth medium of wild-type of transformed strains that produces the APS. This can be undertaken using known protocols for the isolation of molecules of known characteristics.

For example, for APSs which contain multiple benzene rings (pyrrolnitrin and soraphen) cultures are grown for 24 h in 10 ml L broth at an appropriate temperature and then extracted with an equal volume of ethyl acetate. The organic phase is recovered, allowed to evaporated under vacuum and the residue dissolved in 20 l of methanol.

In the case of pyrrolnitrin a further procedure has been used successfully for the extraction of the active antipathogenic compound from the growth medium of the transformed strain producing this antibiotic. This is accomplished by extraction of the medium with 80% acetone followed by removal of the acetone by evaporation and a second extraction with diethyl ether. The diethyl ether is removed by evaporation and the dried extract is resuspended in a small volume of water. Small aliquots of the antibiotic extract applied to small sterile filter paper discs placed on an agar plate will inhibit the growth of *Rhizoctonia solani*, indicating the presence of the active antibiotic compound.

A preferred method for phenazine isolation is described by Thomashow *et al.* (Appl Environ Microbiol 56: 908-912 (1990)). This involves acidifying cultures to pH 2.0 with HCl and extraction with benzene. Benzene fractions are dehydrated with Na_2SO_4 and evaporated to dryness. The residue is redissolved in aqueous 5% $NaHCO_3$, reextracted with an equal volume of benzene, acidified, partitioned into benzene and redried.

For peptide antibiotics (which are typically hydrophobic) extraction techniques using butanol, methanol, chloroform or hexane are suitable. In the case of gramicidin, isolation can be carried out according to the procedure described by Gause & Brazhnikova (Lancet 247: 715 (1944)). For epidermin, the procedure described by Allgaier *et al.* for epidermin (Eur. Ju. Biochem. 160: 9-22 (1986)) is suitable and involves butanol extraction, and dissolving in methanol and diethyl ether. For many APSs (*e.g.* pyrrolnitrin, gramicidin, phenazine) appropriate techniques are provided in the Merck Index (Merck & Co., Rahway, NJ (1989)).

P. Formulation and Use of Isolated Antibiotics

Antifungal formulations can be made using active ingredients which comprise either the isolated APSs or alternatively suspensions or concentrates of cells which produce them. Formulations can be made in liquid or solid form.

Example 58: **Liquid Formulation of Antifungal Compositions**

In the following examples, percentages of composition are given by weight:

1. **Emulsifiable concentrates:**	**a**	**b**	**c**
Active ingredient	20%	40%	50%
Calcium dodecylbenzenesulfonate	5%	8%	6%
Castor oil polyethlene glycol ether (36 moles of ethylene oxide)	5%	-	-
Tributylphenol polyethylene glyco ether (30 moles of ethylene oxide)	-	12%	4%
Cyclohexanone	-	15%	20%
Xylene mixture	70%	25 %	20%

Emulsions of any required concentration can be produced from such concentrates by dilution with water.

2. **Solutions:**	**a**	**b**	**c**	**d**
Active ingredient	80%	10%	5%	95%
Ethylene glycol monomethyl ether	20%	-	-	-
Polyethylene glycol 400	-	70%	-	-
N-methyl-2-pyrrolidone	-	20 %	-	-
Epoxidised coconut oil	-	-	1%	5%
Petroleum distillate (boiling range 160-190°)	-	-	94%	-

These solutions are suitable for application in the form of microdrops.

3. **Granulates:**	**a**	**b**
Active ingredient	5%	10%
Kaolin	94%	-
Highly dispersed silicic acid	1%	-
Attapulgit	-	90%

The active ingredient is dissolved in methylene chloride, the solution is sprayed onto the carrier, and the solvent is subsequently evaporated off in vacuo.

4. **Dusts:**	**a**	**b**
Active ingredient	2%	5%
Highly dispersed silicic acid	1%	5%
Talcum	97%	-
Kaolin	-	90%

Ready-to-use dusts are obtained by intimately mixing the carriers with the active ingredient.

Example 59: **Solid Formulation of Antifungal Compositions**
In the following examples, percentages of compositions are by weight.

1. **Wettable powders:**	**a**	**b**	**c**
Active ingredient	20%	60%	75%
Sodium lignosulfonate	5%	5%	-
Sodium lauryl sulfate	3%	-	5%
Sodium diisobutylnaphthalene sulfonate	-	6%	10 %
Octylphenol polyethylene glycol ether (7-8 moles of ethylene oxide)	-	2%	-
Highly dispersed silicic acid	5%	27%	10%
Kaolin	67%	-	-

The active ingredient is thoroughly mixed with the adjuvants and the mixture is thoroughly ground in a suitable mill, affording wettable powders which can be diluted with water to give suspensions of the desired concentrations.

2. **Emulsifiable concentrate:**	
Active ingredient	10%
Octylphenol polyethylene glycol ether (4-5 moles of ethylene oxide)	3%
Calcium dodecylbenzenesulfonate	3%
Castor oil polyglycol ether (36 moles of ethylene oxide)	4%
Cyclohexanone	30%
Xylene mixture	50%

Emulsions of any required concentration can be obtained from this concentrate by dilution with water.

3. **Dusts:**	a	b
Active ingredient	5%	8%
Talcum	95%	-
Kaolin	-	92%

Ready-to-use dusts are obtained by mixing the active ingredient with the carriers, and grinding the mixture in a suitable mill.

4. **Extruder granulate:**	
Active ingredient	10%
Sodium lignosulfonate	2%
Carboxymethylcellulose	1%
Kaolin	87%

The active ingredient is mixed and ground with the adjuvants, and the mixture is subsequently moistened with water. The mixture is extruded and then dried in a stream of air.

5. **Coated granulate:**	
Active ingredient	3%
Polyethylene glycol 200	3%
Kaolin	94%

The finely ground active ingredient is uniformly applied, in a mixer, to the kaolin moistened with polyethylene glycol. Non-dusty coated granulates are obtained in this manner.

6. **Suspension concentrate:**	
Active ingredient	40%
Ethylene glycol	10%
Nonylphenol polyethylene glycol (15 moles of ethylene oxide)	6%

Sodium lignosulfonate	10%
Carboxymethylcellulose	1%
37 % aqueous formaldehyde solution	0.2%
Silicone oil in 75 % aqueous emulsion	0.8%
Water	32%

The finely ground active ingredient is intimately mixed with the adjuvants, giving a suspension concentrate from which suspensions of any desire concentration can be obtained by dilution with water.

While the present invention has been described with reference to specific embodiments thereof, it will be appreciated that numerous variations, modifications, and embodiments are possible, and accordingly, all such variations, modifications and embodiments are to be regarded as being within the spirit and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: CIBA-GEIGY AG
(B) STREET: Klybeckstr. 141
(C) CITY: Basel
(E) COUNTRY: Switzerland
(F) POSTAL CODE (ZIP): 4002
(G) TELEPHONE: +41 61 69 11 11
(H) TELEFAX: + 41 61 696 79 76
(I) TELEX: 962 991

(ii) TITLE OF INVENTION: Genes for the synthesis of antipathogenic substances

(iii) NUMBER OF SEQUENCES: 22

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7000 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(B) STRAIN: single

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 357..2039
(D) OTHER INFORMATION: /label= ORF1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2249..3076
(D) OTHER INFORMATION: /label= ORF2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3166..4869
(D) OTHER INFORMATION: /label= ORF3

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4894..5985
(D) OTHER INFORMATION: /label= ORF4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GAATTCCGAC AACGCCGAAG AAGCGCGGAA CCGCTGAAAG AGGAGCAGGA ACTGGAGCAA        60

ACGCTGTCCC AGGTGATCGA CAGCCTGCCA CTGCGCATCG AGGGCCGATG AACAGCATTG       120

GCAAAAGCTG GCGGTGCGCA GTGCGCGAGT GATCCGATCA TTTTTGATCG GCTCGCCTCT       180

TCAAAATCGG CGGTGGATGA AGTCGACGGC GGACTGATCA GGCGCAAAAG AACATGCGCC       240

AAAACCTTCT TTTATAGCGA ATACCTTTGC ACTTCAGAAT GTTAATTCGG AAACGGAATT       300

TGCATCGCTT TTCCGGCAGT CTAGAGTCTC TAACAGCACA TTGATGTGCC TCTTGC           356

ATG GAT GCA CGA AGA CTG GCG GCC TCC CCT CGT CAC AGG CGG CCC GCC         404
Met Asp Ala Arg Arg Leu Ala Ala Ser Pro Arg His Arg Arg Pro Ala
  1               5                  10                  15

TTT GAC ACA AGG AGT GTT ATG AAC AAG CCG ATC AAG AAT ATC GTC ATC         452
Phe Asp Thr Arg Ser Val Met Asn Lys Pro Ile Lys Asn Ile Val Ile
                 20                  25                  30

GTG GGC GGC GGT ACT GCG GGC TGG ATG GCC GCC TCG TAC CTC GTC CGG         500
Val Gly Gly Gly Thr Ala Gly Trp Met Ala Ala Ser Tyr Leu Val Arg
         35                  40                  45

GCC CTC CAA CAG CAG GCG AAC ATT ACG CTC ATC GAA TCT GCG GCG ATC         548
Ala Leu Gln Gln Gln Ala Asn Ile Thr Leu Ile Glu Ser Ala Ala Ile
     50                  55                  60

CCT CGG ATC GGC GTG GGC GAA GCG ACC ATC CCA AGT TTG CAG AAG GTG         596
Pro Arg Ile Gly Val Gly Glu Ala Thr Ile Pro Ser Leu Gln Lys Val
 65                  70                  75                  80

TTC TTC GAT TTC CTC GGG ATA CCG GAG CGG GAA TGG ATG CCC CAA GTG         644
Phe Phe Asp Phe Leu Gly Ile Pro Glu Arg Glu Trp Met Pro Gln Val
                 85                  90                  95

AAC GGC GCG TTC AAG GCC GCG ATC AAG TTC GTG AAT TGG AGA AAG TCT         692
Asn Gly Ala Phe Lys Ala Ala Ile Lys Phe Val Asn Trp Arg Lys Ser
                100                 105                 110

CCC GAC CCC TCG CGC GAC GAT CAC TTC TAC CAT TTG TTC GGC AAC GTG         740
Pro Asp Pro Ser Arg Asp Asp His Phe Tyr His Leu Phe Gly Asn Val
```

```
        115                 120                 125

CCG AAC TGC GAC GGC GTG CCG CTT ACC CAC TAC TGG CTG CGC AAG CGC      788
Pro Asn Cys Asp Gly Val Pro Leu Thr His Tyr Trp Leu Arg Lys Arg
    130                 135                 140

GAA CAG GGC TTC CAG CAG CCG ATG GAG TAC GCG TGC TAC CCG CAG CCC      836
Glu Gln Gly Phe Gln Gln Pro Met Glu Tyr Ala Cys Tyr Pro Gln Pro
145                 150                 155                 160

GGG GCA CTC GAC GGC AAG CTG GCA CCG TGC CTG TCC GAC GGC ACC CGC      884
Gly Ala Leu Asp Gly Lys Leu Ala Pro Cys Leu Ser Asp Gly Thr Arg
                165                 170                 175

CAG ATG TCC CAC GCG TGG CAC TTC GAC GCG CAC CTG GTG GCC GAC TTC      932
Gln Met Ser His Ala Trp His Phe Asp Ala His Leu Val Ala Asp Phe
            180                 185                 190

TTG AAG CGC TGG GCC GTC GAG CGC GGG GTG AAC CGC GTG GTC GAT GAG      980
Leu Lys Arg Trp Ala Val Glu Arg Gly Val Asn Arg Val Val Asp Glu
        195                 200                 205

GTG GTG GAC GTT CGC CTG AAC AAC CGC GGC TAC ATC TCC AAC CTG CTC     1028
Val Val Asp Val Arg Leu Asn Asn Arg Gly Tyr Ile Ser Asn Leu Leu
    210                 215                 220

ACC AAG GAG GGG CGG ACG CTG GAG GCG GAC CTG TTC ATC GAC TGC TCC     1076
Thr Lys Glu Gly Arg Thr Leu Glu Ala Asp Leu Phe Ile Asp Cys Ser
225                 230                 235                 240

GGC ATG CGG GGG CTC CTG ATC AAT CAG GCG CTG AAG GAA CCC TTC ATC     1124
Gly Met Arg Gly Leu Leu Ile Asn Gln Ala Leu Lys Glu Pro Phe Ile
                245                 250                 255

GAC ATG TCC GAC TAC CTG CTG TGC GAC AGC GCG GTC GCC AGC GCC GTG     1172
Asp Met Ser Asp Tyr Leu Leu Cys Asp Ser Ala Val Ala Ser Ala Val
            260                 265                 270

CCC AAC GAC GAC GCG CGC GAT GGG GTC GAG CCG TAC ACC TCC TCG ATC     1220
Pro Asn Asp Asp Ala Arg Asp Gly Val Glu Pro Tyr Thr Ser Ser Ile
        275                 280                 285

GCC ATG AAC TCG GGA TGG ACC TGG AAG ATT CCG ATG CTG GGC CGG TTC     1268
Ala Met Asn Ser Gly Trp Thr Trp Lys Ile Pro Met Leu Gly Arg Phe
    290                 295                 300

GGC AGC GGC TAC GTC TTC TCG AGC CAT TTC ACC TCG CGC GAC CAG GCC     1316
Gly Ser Gly Tyr Val Phe Ser Ser His Phe Thr Ser Arg Asp Gln Ala
305                 310                 315                 320

ACC GCC GAC TTC CTC AAA CTC TGG GGC CTC TCG GAC AAT CAG CCG CTC     1364
Thr Ala Asp Phe Leu Lys Leu Trp Gly Leu Ser Asp Asn Gln Pro Leu
                325                 330                 335

AAC CAG ATC AAG TTC CGG GTC GGG CGC AAC AAG CGG GCG TGG GTC AAC     1412
```

Asn Gln Ile Lys Phe Arg Val Gly Arg Asn Lys Arg Ala Trp Val Asn
340 345 350

AAC TGC GTC TCG ATC GGG CTG TCG TCG TGC TTT CTG GAG CCC CTG GAA 1460
Asn Cys Val Ser Ile Gly Leu Ser Ser Cys Phe Leu Glu Pro Leu Glu
355 360 365

TCG ACG GGG ATC TAC TTC ATC TAC GCG GCG CTT TAC CAG CTC GTG AAG 1508
Ser Thr Gly Ile Tyr Phe Ile Tyr Ala Ala Leu Tyr Gln Leu Val Lys
370 375 380

CAC TTC CCC GAC ACC TCG TTC GAC CCG CGG CTG AGC GAC GCT TTC AAC 1556
His Phe Pro Asp Thr Ser Phe Asp Pro Arg Leu Ser Asp Ala Phe Asn
385 390 395 400

GCC GAG ATC GTC CAC ATG TTC GAC GAC TGC CGG GAT TTC GTC CAA GCG 1604
Ala Glu Ile Val His Met Phe Asp Asp Cys Arg Asp Phe Val Gln Ala
405 410 415

CAC TAT TTC ACC ACG TCG CGC GAT GAC ACG CCG TTC TGG CTC GCG AAC 1652
His Tyr Phe Thr Thr Ser Arg Asp Asp Thr Pro Phe Trp Leu Ala Asn
420 425 430

CGG CAC GAC CTG CGG CTC TCG GAC GCC ATC AAA GAG AAG GTT CAG CGC 1700
Arg His Asp Leu Arg Leu Ser Asp Ala Ile Lys Glu Lys Val Gln Arg
435 440 445

TAC AAG GCG GGG CTG CCG CTG ACC ACC ACG TCG TTC GAC GAT TCC ACG 1748
Tyr Lys Ala Gly Leu Pro Leu Thr Thr Thr Ser Phe Asp Asp Ser Thr
450 455 460

TAC TAC GAG ACC TTC GAC TAC GAA TTC AAG AAT TTC TGG TTG AAC GGC 1796
Tyr Tyr Glu Thr Phe Asp Tyr Glu Phe Lys Asn Phe Trp Leu Asn Gly
465 470 475 480

AAC TAC TAC TGC ATC TTT GCC GGC TTG GGC ATG CTG CCC GAC CGG TCG 1844
Asn Tyr Tyr Cys Ile Phe Ala Gly Leu Gly Met Leu Pro Asp Arg Ser
485 490 495

CTG CCG CTG TTG CAG CAC CGA CCG GAG TCG ATC GAG AAA GCC GAG GCG 1892
Leu Pro Leu Leu Gln His Arg Pro Glu Ser Ile Glu Lys Ala Glu Ala
500 505 510

ATG TTC GCC AGC ATC CGG CGC GAG GCC GAG CGT CTG CGC ACC AGC CTG 1940
Met Phe Ala Ser Ile Arg Arg Glu Ala Glu Arg Leu Arg Thr Ser Leu
515 520 525

CCG ACA AAC TAC GAC TAC CTG CGG TCG CTG CGT GAC GGC GAC GCG GGG 1988
Pro Thr Asn Tyr Asp Tyr Leu Arg Ser Leu Arg Asp Gly Asp Ala Gly
530 535 540

CTG TCG CGC GGC CAG CGT GGG CCG AAG CTC GCA GCG CAG GAA AGC CTG 2036
Leu Ser Arg Gly Gln Arg Gly Pro Lys Leu Ala Ala Gln Glu Ser Leu
545 550 555 560

```
TAGTGGAACG CACCTTGGAC CGGGTAGGCG TATTCGCGGC CACCCACGCT GCCGTGGCGG     2096

CCTGCGATCC GCTGCAGGCG CGCGCGCTCG TTCTGCAACT GCCGGGCCTG AACCGTAACA     2156

AGGACGTGCC CGGTATCGTC GGCCTGCTGC GCGAGTTCCT TCCGGTGCGC GGCCTGCCCT     2216

GCGGCTGGGG TTTCGTCGAA GCCGCCGCCG CG ATG CGG GAC ATC GGG TTC TTC       2269
                                    Met Arg Asp Ile Gly Phe Phe
                                     1               5

CTG GGG TCG CTC AAG CGC CAC GGA CAT GAG CCC GCG GAG GTG GTG CCC      2317
Leu Gly Ser Leu Lys Arg His Gly His Glu Pro Ala Glu Val Val Pro
        10                  15                  20

GGG CTT GAG CCG GTG CTG CTC GAC CTG GCA CGC GCG ACC AAC CTG CCG      2365
Gly Leu Glu Pro Val Leu Leu Asp Leu Ala Arg Ala Thr Asn Leu Pro
    25                  30                  35

CCG CGC GAG ACG CTC CTG CAT GTG ACG GTC TGG AAC CCC ACG GCG GCC      2413
Pro Arg Glu Thr Leu Leu His Val Thr Val Trp Asn Pro Thr Ala Ala
 40                  45                  50                  55

GAC GCG CAG CGC AGC TAC ACC GGG CTG CCC GAC GAA GCG CAC CTG CTC      2461
Asp Ala Gln Arg Ser Tyr Thr Gly Leu Pro Asp Glu Ala His Leu Leu
                60                  65                  70

GAG AGC GTG CGC ATC TCG ATG GCG GCC CTC GAG GCG GCC ATC GCG TTG      2509
Glu Ser Val Arg Ile Ser Met Ala Ala Leu Glu Ala Ala Ile Ala Leu
            75                  80                  85

ACC GTC GAG CTG TTC GAT GTG TCC CTG CGG TCG CCC GAG TTC GCG CAA      2557
Thr Val Glu Leu Phe Asp Val Ser Leu Arg Ser Pro Glu Phe Ala Gln
        90                  95                 100

AGG TGC GAC GAG CTG GAA GCC TAT CTG CAG AAA ATG GTC GAA TCG ATC      2605
Arg Cys Asp Glu Leu Glu Ala Tyr Leu Gln Lys Met Val Glu Ser Ile
    105                 110                 115

GTC TAC GCG TAC CGC TTC ATC TCG CCG CAG GTC TTC TAC GAT GAG CTG      2653
Val Tyr Ala Tyr Arg Phe Ile Ser Pro Gln Val Phe Tyr Asp Glu Leu
120                 125                 130                 135

CGC CCC TTC TAC GAA CCG ATT CGA GTC GGG GGC CAG AGC TAC CTC GGC      2701
Arg Pro Phe Tyr Glu Pro Ile Arg Val Gly Gly Gln Ser Tyr Leu Gly
                140                 145                 150

CCC GGT GCC GTA GAG ATG CCC CTC TTC GTG CTG GAG CAC GTC CTC TGG      2749
Pro Gly Ala Val Glu Met Pro Leu Phe Val Leu Glu His Val Leu Trp
            155                 160                 165

GGC TCG CAA TCG GAC GAC CAA ACT TAT CGA GAA TTC AAA GAG ACG TAC      2797
Gly Ser Gln Ser Asp Asp Gln Thr Tyr Arg Glu Phe Lys Glu Thr Tyr
        170                 175                 180

CTG CCC TAT GTG CTT CCC GCG TAC AGG GCG GTC TAC GCT CGG TTC TCC      2845
```

```
Leu Pro Tyr Val Leu Pro Ala Tyr Arg Ala Val Tyr Ala Arg Phe Ser
    185                 190                 195

GGG GAG CCG GCG CTC ATC GAC CGC GCG CTC GAC GAG GCG CGA GCG GTC      2893
Gly Glu Pro Ala Leu Ile Asp Arg Ala Leu Asp Glu Ala Arg Ala Val
200                 205                 210                 215

GGT ACG CGG GAC GAG CAC GTC CGG GCT GGG CTG ACA GCC CTC GAG CGG      2941
Gly Thr Arg Asp Glu His Val Arg Ala Gly Leu Thr Ala Leu Glu Arg
                220                 225                 230

GTC TTC AAG GTC CTG CTG CGC TTC CGG GCG CCT CAC CTC AAA TTG GCG      2989
Val Phe Lys Val Leu Leu Arg Phe Arg Ala Pro His Leu Lys Leu Ala
            235                 240                 245

GAG CGG GCG TAC GAA GTC GGG CAA AGC GGC CCG AAA TCG GCA GCG GGG      3037
Glu Arg Ala Tyr Glu Val Gly Gln Ser Gly Pro Lys Ser Ala Ala Gly
        250                 255                 260

GGT ACG CGC CCA GCA TGC TCG GTG AGC TGC TCA CGC TGACGTATGC           3083
Gly Thr Arg Pro Ala Cys Ser Val Ser Cys Ser Arg
    265                 270                 275

CGCGCGGTCC CGCGTCCGCG CCGCGCTCGA CGAATCCTGA TGCGCGCGAC CCAGTGTTAT    3143

CTCACAAGGA GAGTTTGCCC CC ATG ACT CAG AAG AGC CCC GCG AAC GAA CAC     3195
                         Met Thr Gln Lys Ser Pro Ala Asn Glu His
                          1               5                  10

GAT AGC AAT CAC TTC GAC GTA ATC ATC CTC GGC TCG GGC ATG TCC GGC      3243
Asp Ser Asn His Phe Asp Val Ile Ile Leu Gly Ser Gly Met Ser Gly
                15                  20                  25

ACC CAG ATG GGG GCC ATC TTG GCC AAA CAA CAG TTT CGC GTG CTG ATC      3291
Thr Gln Met Gly Ala Ile Leu Ala Lys Gln Gln Phe Arg Val Leu Ile
            30                  35                  40

ATC GAG GAG TCG TCG CAC CCG CGG TTC ACG ATC GGC GAA TCG TCG ATC      3339
Ile Glu Glu Ser Ser His Pro Arg Phe Thr Ile Gly Glu Ser Ser Ile
        45                  50                  55

CCC GAG ACG TCT CTT ATG AAC CGC ATC ATC GCT GAT CGC TAC GGC ATT      3387
Pro Glu Thr Ser Leu Met Asn Arg Ile Ile Ala Asp Arg Tyr Gly Ile
    60                  65                  70

CCG GAG CTC GAC CAC ATC ACG TCG TTT TAT TCG ACG CAA CGT TAC GTC      3435
Pro Glu Leu Asp His Ile Thr Ser Phe Tyr Ser Thr Gln Arg Tyr Val
75                  80                  85                  90

GCG TCG AGC ACG GGC ATT AAG CGC AAC TTC GGC TTC GTG TTC CAC AAG      3483
Ala Ser Ser Thr Gly Ile Lys Arg Asn Phe Gly Phe Val Phe His Lys
                95                 100                 105

CCC GGC CAG GAG CAC GAC CCG AAG GAG TTC ACC CAG TGC GTC ATT CCC      3531
Pro Gly Gln Glu His Asp Pro Lys Glu Phe Thr Gln Cys Val Ile Pro
```

110 115 120

GAG CTG CCG TGG GGG CCG GAG AGC CAT TAT TAC CGG CAA GAC GTC GAC 3579
Glu Leu Pro Trp Gly Pro Glu Ser His Tyr Tyr Arg Gln Asp Val Asp
125 130 135

GCC TAC TTG TTG CAA GCC GCC ATT AAA TAC GGC TGC AAG GTC CAC CAG 3627
Ala Tyr Leu Leu Gln Ala Ala Ile Lys Tyr Gly Cys Lys Val His Gln
140 145 150

AAA ACT ACC GTG ACC GAA TAC CAC GCC GAT AAA GAC GGC GTC GCG GTG 3675
Lys Thr Thr Val Thr Glu Tyr His Ala Asp Lys Asp Gly Val Ala Val
155 160 165 170

ACC ACC GCC CAG GGC GAA CGG TTC ACC GGC CGG TAC ATG ATC GAC TGC 3723
Thr Thr Ala Gln Gly Glu Arg Phe Thr Gly Arg Tyr Met Ile Asp Cys
175 180 185

GGA GGA CCT CGC GCG CCG CTC GCG ACC AAG TTC AAG CTC CGC GAA GAA 3771
Gly Gly Pro Arg Ala Pro Leu Ala Thr Lys Phe Lys Leu Arg Glu Glu
190 195 200

CCG TGT CGC TTC AAG ACG CAC TCG CGC AGC CTC TAC ACG CAC ATG CTC 3819
Pro Cys Arg Phe Lys Thr His Ser Arg Ser Leu Tyr Thr His Met Leu
205 210 215

GGG GTC AAG CCG TTC GAC GAC ATC TTC AAG GTC AAG GGG CAG CGC TGG 3867
Gly Val Lys Pro Phe Asp Asp Ile Phe Lys Val Lys Gly Gln Arg Trp
220 225 230

CGC TGG CAC GAG GGG ACC TTG CAC CAC ATG TTC GAG GGC GGC TGG CTC 3915
Arg Trp His Glu Gly Thr Leu His His Met Phe Glu Gly Gly Trp Leu
235 240 245 250

TGG GTG ATT CCG TTC AAC AAC CAC CCG CGG TCG ACC AAC AAC CTG GTG 3963
Trp Val Ile Pro Phe Asn Asn His Pro Arg Ser Thr Asn Asn Leu Val
255 260 265

AGC GTC GGC CTG CAG CTC GAC CCG CGT GTC TAC CCG AAA ACC GAC ATC 4011
Ser Val Gly Leu Gln Leu Asp Pro Arg Val Tyr Pro Lys Thr Asp Ile
270 275 280

TCC GCA CAG CAG GAA TTC GAT GAG TTC CTC GCG CGG TTC CCG AGC ATC 4059
Ser Ala Gln Gln Glu Phe Asp Glu Phe Leu Ala Arg Phe Pro Ser Ile
285 290 295

GGG GCT CAG TTC CGG GAC GCC GTG CCG GTG CGC GAC TGG GTC AAG ACC 4107
Gly Ala Gln Phe Arg Asp Ala Val Pro Val Arg Asp Trp Val Lys Thr
300 305 310

GAC CGC CTG CAA TTC TCG TCG AAC GCC TGC GTC GGC GAC CGC TAC TGC 4155
Asp Arg Leu Gln Phe Ser Ser Asn Ala Cys Val Gly Asp Arg Tyr Cys
315 320 325 330

CTG ATG CTG CAC GCG AAC GGC TTC ATC GAC CCG CTC TTC TCC GGG GGG 4203

Leu Met Leu His Ala Asn Gly Phe Ile Asp Pro Leu Phe Ser Arg Gly
335 340 345

CTG GAA AAC ACC GCG GTG ACC ATC CAC GCG CTC GCG GCG CGC CTC ATC 4251
Leu Glu Asn Thr Ala Val Thr Ile His Ala Leu Ala Ala Arg Leu Ile
350 355 360

AAG GCG CTG CGC GAC GAC GAC TTC TCC CCC GAG CGC TTC GAG TAC ATC 4299
Lys Ala Leu Arg Asp Asp Asp Phe Ser Pro Glu Arg Phe Glu Tyr Ile
365 370 375

GAG CGC CTG CAG CAA AAG CTT TTG GAC CAC AAC GAC GAC TTC GTC AGC 4347
Glu Arg Leu Gln Gln Lys Leu Leu Asp His Asn Asp Asp Phe Val Ser
380 385 390

TGC TGC TAC ACG GCG TTC TCG GAC TTC CGC CTA TGG GAC GCG TTC CAC 4395
Cys Cys Tyr Thr Ala Phe Ser Asp Phe Arg Leu Trp Asp Ala Phe His
395 400 405 410

AGG CTG TGG GCG GTC GGC ACC ATC CTC GGG CAG TTC CGG CTC GTG CAG 4443
Arg Leu Trp Ala Val Gly Thr Ile Leu Gly Gln Phe Arg Leu Val Gln
415 420 425

GCC CAC GCG AGG TTC CGC GCG TCG CGC AAC GAG GGC GAC CTC GAT CAC 4491
Ala His Ala Arg Phe Arg Ala Ser Arg Asn Glu Gly Asp Leu Asp His
430 435 440

CTC GAC AAC GAC CCT CCG TAT CTC GGA TAC CTG TGC GCG GAC ATG GAG 4539
Leu Asp Asn Asp Pro Pro Tyr Leu Gly Tyr Leu Cys Ala Asp Met Glu
445 450 455

GAG TAC TAC CAG TTG TTC AAC GAC GCC AAA GCC GAG GTC GAG GCC GTG 4587
Glu Tyr Tyr Gln Leu Phe Asn Asp Ala Lys Ala Glu Val Glu Ala Val
460 465 470

AGT GCC GGG CGC AAG CCG GCC GAT GAG GCC GCG GCG CGG ATT CAC GCC 4635
Ser Ala Gly Arg Lys Pro Ala Asp Glu Ala Ala Ala Arg Ile His Ala
475 480 485 490

CTC ATT GAC GAA CGA GAC TTC GCC AAG CCG ATG TTC GGC TTC GGG TAC 4683
Leu Ile Asp Glu Arg Asp Phe Ala Lys Pro Met Phe Gly Phe Gly Tyr
495 500 505

TGC ATC ACC GGG GAC AAG CCG CAG CTC AAC AAC TCG AAG TAC AGC CTG 4731
Cys Ile Thr Gly Asp Lys Pro Gln Leu Asn Asn Ser Lys Tyr Ser Leu
510 515 520

CTG CCG GCG ATG CGG CTG ATG TAC TGG ACG CAA ACC CGC GCG CCG GCA 4779
Leu Pro Ala Met Arg Leu Met Tyr Trp Thr Gln Thr Arg Ala Pro Ala
525 530 535

GAG GTG AAA AAG TAC TTC GAC TAC AAC CCG ATG TTC GCG CTG CTC AAG 4827
Glu Val Lys Lys Tyr Phe Asp Tyr Asn Pro Met Phe Ala Leu Leu Lys
540 545 550

```
GCG TAC ATC ACG ACC CGC ATC GGC CTG GCG CTG AAG AAG TAGCCGCTCG      4876
Ala Tyr Ile Thr Thr Arg Ile Gly Leu Ala Leu Lys Lys
555                 560                 565

ACGACGACAT AAAAACG ATG AAC GAC ATT CAA TTG GAT CAA GCG AGC GTC      4926
                   Met Asn Asp Ile Gln Leu Asp Gln Ala Ser Val
                    1               5                   10

AAG AAG CGT CCC TCG GGC GCG TAC GAC GCA ACC ACG CGC CTG GCC GCG     4974
Lys Lys Arg Pro Ser Gly Ala Tyr Asp Ala Thr Thr Arg Leu Ala Ala
            15                  20                  25

AGC TGG TAC GTC GCG ATG CGC TCC AAC GAG CTC AAG GAC AAG CCG ACC     5022
Ser Trp Tyr Val Ala Met Arg Ser Asn Glu Leu Lys Asp Lys Pro Thr
        30                  35                  40

GAG TTG ACG CTC TTC GGC CGT CCG TGC GTG GCG TGG CGC GGA GCC ACG     5070
Glu Leu Thr Leu Phe Gly Arg Pro Cys Val Ala Trp Arg Gly Ala Thr
    45                  50                  55

GGG CGG GCC GTG GTG ATG GAC CGC CAC TGC TCG CAC CTG GGC GCG AAC     5118
Gly Arg Ala Val Val Met Asp Arg His Cys Ser His Leu Gly Ala Asn
60                  65                  70                  75

CTG GCT GAC GGG CGG ATC AAG GAC GGG TGC ATC CAG TGC CCG TTT CAC     5166
Leu Ala Asp Gly Arg Ile Lys Asp Gly Cys Ile Gln Cys Pro Phe His
                80                  85                  90

CAC TGG CGG TAC GAC GAA CAG GGC CAG TGC GTT CAC ATC CCC GGC CAT     5214
His Trp Arg Tyr Asp Glu Gln Gly Gln Cys Val His Ile Pro Gly His
            95                  100                 105

AAC CAG GCG GTG CGC CAG CTG GAG CCG GTG CCG CGC GGG GCG CGT CAG     5262
Asn Gln Ala Val Arg Gln Leu Glu Pro Val Pro Arg Gly Ala Arg Gln
        110                 115                 120

CCG ACG TTG GTC ACC GCC GAG CGA TAC GGC TAC GTG TGG GTC TGG TAC     5310
Pro Thr Leu Val Thr Ala Glu Arg Tyr Gly Tyr Val Trp Val Trp Tyr
    125                 130                 135

GGC TCC CCG CTG CCG CTG CAC CCG CTG CCC GAA ATC TCC GCG GCC GAT     5358
Gly Ser Pro Leu Pro Leu His Pro Leu Pro Glu Ile Ser Ala Ala Asp
140                 145                 150                 155

GTC GAC AAC GGC GAC TTT ATG CAC CTG CAC TTC GCG TTC GAG ACG ACC     5406
Val Asp Asn Gly Asp Phe Met His Leu His Phe Ala Phe Glu Thr Thr
                160                 165                 170

ACG GCG GTC TTG CGG ATC GTC GAG AAC TTC TAC GAC GCG CAG CAC GCA     5454
Thr Ala Val Leu Arg Ile Val Glu Asn Phe Tyr Asp Ala Gln His Ala
            175                 180                 185

ACC CCG GTG CAC GCA CTC CCG ATC TCG GCC TTC GAA CTC AAG CTC TTC     5502
Thr Pro Val His Ala Leu Pro Ile Ser Ala Phe Glu Leu Lys Leu Phe
        190                 195                 200
```

GAC GAT TGG CGC CAG TGG CCG GAG GTT GAG TCG CTG GCC CTG GCG GGC 5550
Asp Asp Trp Arg Gln Trp Pro Glu Val Glu Ser Leu Ala Leu Ala Gly
205 210 215

GCG TGG TTC GGT GCC GGG ATC GAC TTC ACC GTG GAC CGG TAC TTC GGC 5598
Ala Trp Phe Gly Ala Gly Ile Asp Phe Thr Val Asp Arg Tyr Phe Gly
220 225 230 235

CCC CTC GGC ATG CTG TCA CGC GCG CTC GGC CTG AAC ATG TCG CAG ATG 5646
Pro Leu Gly Met Leu Ser Arg Ala Leu Gly Leu Asn Met Ser Gln Met
240 245 250

AAC CTG CAC TTC GAT GGC TAC CCC GGC GGG TGC GTC ATG ACC GTC GCC 5694
Asn Leu His Phe Asp Gly Tyr Pro Gly Gly Cys Val Met Thr Val Ala
255 260 265

CTG GAC GGA GAC GTC AAA TAC AAG CTG CTC CAG TGT GTG ACG CCG GTG 5742
Leu Asp Gly Asp Val Lys Tyr Lys Leu Leu Gln Cys Val Thr Pro Val
270 275 280

AGC GAA GGC AAG AAC GTC ATG CAC ATG CTC ATC TCG ATC AAG AAG GTG 5790
Ser Glu Gly Lys Asn Val Met His Met Leu Ile Ser Ile Lys Lys Val
285 290 295

GGC GGC ATC CTG CTC CGC GCG ACC GAC TTC GTG CTG TTC GGG CTG CAG 5838
Gly Gly Ile Leu Leu Arg Ala Thr Asp Phe Val Leu Phe Gly Leu Gln
300 305 310 315

ACC AGG CAG GCC GCG GGG TAC GAC GTC AAA ATC TGG AAC GGA ATG AAG 5886
Thr Arg Gln Ala Ala Gly Tyr Asp Val Lys Ile Trp Asn Gly Met Lys
320 325 330

CCG GAC GGC GGC GGC GCG TAC AGC AAG TAC GAC AAG CTC GTG CTC AAG 5934
Pro Asp Gly Gly Gly Ala Tyr Ser Lys Tyr Asp Lys Leu Val Leu Lys
335 340 345

TAC CGG GCG TTC TAT CGA GGC TGG GTC GAC CGC GTC GCA AGT GAG CGG 5982
Tyr Arg Ala Phe Tyr Arg Gly Trp Val Asp Arg Val Ala Ser Glu Arg
350 355 360

TGATGCGTGA AGCCGAGCCG CTCTCGACCG CGTCGCTGCG CCAGGCGCTC GCGAACCTGG 6042

CGAGCGGCGT GACGATCACG GCCTACGGCG CGCCGGGCCC GCTTGGGCTC GCGGCCACCA 6102

GCTTCGTGTC GGAGTCGCTC TTTGCGAGGT ATTCATGACT ATCTGGCTGT TGCAACTCGT 6162

GCTGGTGATC GCGCTCTGCA ACGTCTGCGG CCGCATTGCC GAACGGCTCG GCCAGTGCGC 6222

GGTCATCGGC GAGATCGCGG CCGGTTTGCT GTTGGGGCCG TCGCTGTTCG GCGTGATCGC 6282

ACCGAGTTTC TACGACCTGT TGTTCGGCCC CCAGGTGCTG TCAGCGATGG CGCAAGTCAG 6342

CGAAGTCGGC CTGGTACTGC TGATGTTCCA GGTCGGCCTG CATATGGAGT TGGGCGAGAC 6402

GCTGCGCGAC AAGCGCTGGC GCATGCCCGT CGCGATCGCA GCGGGCGGGC TCGTCGCACC 6462

GGCCGCGATC GGCATGATCG TCGCCATCGT TTCGAAAGGC ACGCTCGCCA GCGACGCGCC 6522

GGCGCTGCCC TATGTGCTCT TCTGCGGTGT CGCACTTGCG GTATCGGCGG TGCCGGTGAT 6582

GGCGCGCATC ATCGACGACC TGGAGCTCAG CGCCATGGTG GGCGCGCGGC ACGCAATGTC 6642

TGCCGCGATG CTGACGGATG CGCTCGGATG GATGCTGCTT GCAACGATTG CCTCGCTATC 6702

GAGCGGGCCC GGCTGGGCAT TTGCGCGCAT GCTCGTCAGC CTGCTCGCGT ATCTGGTGCT 6762

GTGCGCGCTG CTGGTGCGCT TCGTGGTTCG ACCGACCCTT GCGCGGCTCG CGTCGACCGC 6822

GCATGCGACG CGCGACCGCT TGGCCGTGTT GTTCTGCTTC GTAATGTTGT CGGCACTCGC 6882

GACGTCGCTG ATCGGATTCC ATAGCGCTTT TGGCGCACTT GCCGCGGCGC TGTTCGTGCG 6942

CCGGGTGCCC GGCGTCGCGA AGGAGTGGCG CGACAACGTC GAAGGTTTCG TCAAGCTT 7000

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 560 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Asp Ala Arg Arg Leu Ala Ala Ser Pro Arg His Arg Arg Pro Ala
1 5 10 15

Phe Asp Thr Arg Ser Val Met Asn Lys Pro Ile Lys Asn Ile Val Ile
20 25 30

Val Gly Gly Gly Thr Ala Gly Trp Met Ala Ala Ser Tyr Leu Val Arg
35 40 45

Ala Leu Gln Gln Gln Ala Asn Ile Thr Leu Ile Glu Ser Ala Ala Ile
50 55 60

Pro Arg Ile Gly Val Gly Glu Ala Thr Ile Pro Ser Leu Gln Lys Val
65 70 75 80

Phe Phe Asp Phe Leu Gly Ile Pro Glu Arg Glu Trp Met Pro Gln Val
85 90 95

Asn Gly Ala Phe Lys Ala Ala Ile Lys Phe Val Asn Trp Arg Lys Ser
100 105 110

Pro Asp Pro Ser Arg Asp Asp His Phe Tyr His Leu Phe Gly Asn Val
115 120 125

Pro Asn Cys Asp Gly Val Pro Leu Thr His Tyr Trp Leu Arg Lys Arg
130 135 140

Glu Gln Gly Phe Gln Gln Pro Met Glu Tyr Ala Cys Tyr Pro Gln Pro
145 150 155 160

Gly Ala Leu Asp Gly Lys Leu Ala Pro Cys Leu Ser Asp Gly Thr Arg
165 170 175

Gln Met Ser His Ala Trp His Phe Asp Ala His Leu Val Ala Asp Phe
180 185 190

Leu Lys Arg Trp Ala Val Glu Arg Gly Val Asn Arg Val Val Asp Glu
195 200 205

Val Val Asp Val Arg Leu Asn Asn Arg Gly Tyr Ile Ser Asn Leu Leu
210 215 220

Thr Lys Glu Gly Arg Thr Leu Glu Ala Asp Leu Phe Ile Asp Cys Ser
225 230 235 240

Gly Met Arg Gly Leu Leu Ile Asn Gln Ala Leu Lys Glu Pro Phe Ile
245 250 255

Asp Met Ser Asp Tyr Leu Leu Cys Asp Ser Ala Val Ala Ser Ala Val
260 265 270

Pro Asn Asp Asp Ala Arg Asp Gly Val Glu Pro Tyr Thr Ser Ser Ile
275 280 285

Ala Met Asn Ser Gly Trp Thr Trp Lys Ile Pro Met Leu Gly Arg Phe
290 295 300

Gly Ser Gly Tyr Val Phe Ser Ser His Phe Thr Ser Arg Asp Gln Ala
305 310 315 320

Thr Ala Asp Phe Leu Lys Leu Trp Gly Leu Ser Asp Asn Gln Pro Leu
325 330 335

Asn Gln Ile Lys Phe Arg Val Gly Arg Asn Lys Arg Ala Trp Val Asn
340 345 350

Asn Cys Val Ser Ile Gly Leu Ser Ser Cys Phe Leu Glu Pro Leu Glu
355 360 365

Ser Thr Gly Ile Tyr Phe Ile Tyr Ala Ala Leu Tyr Gln Leu Val Lys
370 375 380

His Phe Pro Asp Thr Ser Phe Asp Pro Arg Leu Ser Asp Ala Phe Asn
385 390 395 400

Ala Glu Ile Val His Met Phe Asp Asp Cys Arg Asp Phe Val Gln Ala
405 410 415

His Tyr Phe Thr Thr Ser Arg Asp Asp Thr Pro Phe Trp Leu Ala Asn
420 425 430

Arg His Asp Leu Arg Leu Ser Asp Ala Ile Lys Glu Lys Val Gln Arg
435 440 445

Tyr Lys Ala Gly Leu Pro Leu Thr Thr Thr Ser Phe Asp Asp Ser Thr
450 455 460

Tyr Tyr Glu Thr Phe Asp Tyr Glu Phe Lys Asn Phe Trp Leu Asn Gly
465 470 475 480

Asn Tyr Tyr Cys Ile Phe Ala Gly Leu Gly Met Leu Pro Asp Arg Ser
485 490 495

Leu Pro Leu Leu Gln His Arg Pro Glu Ser Ile Glu Lys Ala Glu Ala
500 505 510

Met Phe Ala Ser Ile Arg Arg Glu Ala Glu Arg Leu Arg Thr Ser Leu
515 520 525

Pro Thr Asn Tyr Asp Tyr Leu Arg Ser Leu Arg Asp Gly Asp Ala Gly
530 535 540

Leu Ser Arg Gly Gln Arg Gly Pro Lys Leu Ala Ala Gln Glu Ser Leu
545 550 555 560

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 275 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Arg Asp Ile Gly Phe Phe Leu Gly Ser Leu Lys Arg His Gly His
1 5 10 15

Glu Pro Ala Glu Val Val Pro Gly Leu Glu Pro Val Leu Leu Asp Leu
20 25 30

Ala Arg Ala Thr Asn Leu Pro Pro Arg Glu Thr Leu Leu His Val Thr
35 40 45

Val Trp Asn Pro Thr Ala Ala Asp Ala Gln Arg Ser Tyr Thr Gly Leu
50 55 60

Pro Asp Glu Ala His Leu Leu Glu Ser Val Arg Ile Ser Met Ala Ala
65 70 75 80

Leu Glu Ala Ala Ile Ala Leu Thr Val Glu Leu Phe Asp Val Ser Leu

85 90 95

Arg Ser Pro Glu Phe Ala Gln Arg Cys Asp Glu Leu Glu Ala Tyr Leu
100 105 110

Gln Lys Met Val Glu Ser Ile Val Tyr Ala Tyr Arg Phe Ile Ser Pro
115 120 125

Gln Val Phe Tyr Asp Glu Leu Arg Pro Phe Tyr Glu Pro Ile Arg Val
130 135 140

Gly Gly Gln Ser Tyr Leu Gly Pro Gly Ala Val Glu Met Pro Leu Phe
145 150 155 160

Val Leu Glu His Val Leu Trp Gly Ser Gln Ser Asp Asp Gln Thr Tyr
165 170 175

Arg Glu Phe Lys Glu Thr Tyr Leu Pro Tyr Val Leu Pro Ala Tyr Arg
180 185 190

Ala Val Tyr Ala Arg Phe Ser Gly Glu Pro Ala Leu Ile Asp Arg Ala
195 200 205

Leu Asp Glu Ala Arg Ala Val Gly Thr Arg Asp Glu His Val Arg Ala
210 215 220

Gly Leu Thr Ala Leu Glu Arg Val Phe Lys Val Leu Leu Arg Phe Arg
225 230 235 240

Ala Pro His Leu Lys Leu Ala Glu Arg Ala Tyr Glu Val Gly Gln Ser
245 250 255

Gly Pro Lys Ser Ala Ala Gly Gly Thr Arg Pro Ala Cys Ser Val Ser
260 265 270

Cys Ser Arg
275

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 567 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Thr Gln Lys Ser Pro Ala Asn Glu His Asp Ser Asn His Phe Asp
1 5 10 15

Val Ile Ile Leu Gly Ser Gly Met Ser Gly Thr Gln Met Gly Ala Ile
20 25 30

```
Leu Ala Lys Gln Gln Phe Arg Val Leu Ile Ile Glu Glu Ser Ser His
         35                  40                  45

Pro Arg Phe Thr Ile Gly Glu Ser Ser Ile Pro Glu Thr Ser Leu Met
     50                  55                  60

Asn Arg Ile Ile Ala Asp Arg Tyr Gly Ile Pro Glu Leu Asp His Ile
 65                  70                  75                  80

Thr Ser Phe Tyr Ser Thr Gln Arg Tyr Val Ala Ser Ser Thr Gly Ile
                 85                  90                  95

Lys Arg Asn Phe Gly Phe Val Phe His Lys Pro Gly Gln Glu His Asp
            100                 105                 110

Pro Lys Glu Phe Thr Gln Cys Val Ile Pro Glu Leu Pro Trp Gly Pro
        115                 120                 125

Glu Ser His Tyr Tyr Arg Gln Asp Val Asp Ala Tyr Leu Leu Gln Ala
    130                 135                 140

Ala Ile Lys Tyr Gly Cys Lys Val His Gln Lys Thr Thr Val Thr Glu
145                 150                 155                 160

Tyr His Ala Asp Lys Asp Gly Val Ala Val Thr Thr Ala Gln Gly Glu
                165                 170                 175

Arg Phe Thr Gly Arg Tyr Met Ile Asp Cys Gly Gly Pro Arg Ala Pro
            180                 185                 190

Leu Ala Thr Lys Phe Lys Leu Arg Glu Glu Pro Cys Arg Phe Lys Thr
        195                 200                 205

His Ser Arg Ser Leu Tyr Thr His Met Leu Gly Val Lys Pro Phe Asp
    210                 215                 220

Asp Ile Phe Lys Val Lys Gly Gln Arg Trp Arg Trp His Glu Gly Thr
225                 230                 235                 240

Leu His His Met Phe Glu Gly Gly Trp Leu Trp Val Ile Pro Phe Asn
                245                 250                 255

Asn His Pro Arg Ser Thr Asn Asn Leu Val Ser Val Gly Leu Gln Leu
            260                 265                 270

Asp Pro Arg Val Tyr Pro Lys Thr Asp Ile Ser Ala Gln Gln Glu Phe
        275                 280                 285

Asp Glu Phe Leu Ala Arg Phe Pro Ser Ile Gly Ala Gln Phe Arg Asp
    290                 295                 300

Ala Val Pro Val Arg Asp Trp Val Lys Thr Asp Arg Leu Gln Phe Ser
305                 310                 315                 320
```

```
Ser Asn Ala Cys Val Gly Asp Arg Tyr Cys Leu Met Leu His Ala Asn
                325                 330                 335

Gly Phe Ile Asp Pro Leu Phe Ser Arg Gly Leu Glu Asn Thr Ala Val
            340                 345                 350

Thr Ile His Ala Leu Ala Ala Arg Leu Ile Lys Ala Leu Arg Asp Asp
        355                 360                 365

Asp Phe Ser Pro Glu Arg Phe Glu Tyr Ile Glu Arg Leu Gln Gln Lys
    370                 375                 380

Leu Leu Asp His Asn Asp Asp Phe Val Ser Cys Cys Tyr Thr Ala Phe
385                 390                 395                 400

Ser Asp Phe Arg Leu Trp Asp Ala Phe His Arg Leu Trp Ala Val Gly
                405                 410                 415

Thr Ile Leu Gly Gln Phe Arg Leu Val Gln Ala His Ala Arg Phe Arg
            420                 425                 430

Ala Ser Arg Asn Glu Gly Asp Leu Asp His Leu Asp Asn Asp Pro Pro
        435                 440                 445

Tyr Leu Gly Tyr Leu Cys Ala Asp Met Glu Glu Tyr Tyr Gln Leu Phe
    450                 455                 460

Asn Asp Ala Lys Ala Glu Val Glu Ala Val Ser Ala Gly Arg Lys Pro
465                 470                 475                 480

Ala Asp Glu Ala Ala Ala Arg Ile His Ala Leu Ile Asp Glu Arg Asp
                485                 490                 495

Phe Ala Lys Pro Met Phe Gly Phe Gly Tyr Cys Ile Thr Gly Asp Lys
            500                 505                 510

Pro Gln Leu Asn Asn Ser Lys Tyr Ser Leu Leu Pro Ala Met Arg Leu
        515                 520                 525

Met Tyr Trp Thr Gln Thr Arg Ala Pro Ala Glu Val Lys Lys Tyr Phe
    530                 535                 540

Asp Tyr Asn Pro Met Phe Ala Leu Leu Lys Ala Tyr Ile Thr Thr Arg
545                 550                 555                 560

Ile Gly Leu Ala Leu Lys Lys
                565
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 363 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Met Asn Asp Ile Gln Leu Asp Gln Ala Ser Val Lys Lys Arg Pro Ser
1               5                   10                  15

Gly Ala Tyr Asp Ala Thr Thr Arg Leu Ala Ala Ser Trp Tyr Val Ala
            20                  25                  30

Met Arg Ser Asn Glu Leu Lys Asp Lys Pro Thr Glu Leu Thr Leu Phe
        35                  40                  45

Gly Arg Pro Cys Val Ala Trp Arg Gly Ala Thr Gly Arg Ala Val Val
    50                  55                  60

Met Asp Arg His Cys Ser His Leu Gly Ala Asn Leu Ala Asp Gly Arg
65                  70                  75                  80

Ile Lys Asp Gly Cys Ile Gln Cys Pro Phe His His Trp Arg Tyr Asp
                85                  90                  95

Glu Gln Gly Gln Cys Val His Ile Pro Gly His Asn Gln Ala Val Arg
            100                 105                 110

Gln Leu Glu Pro Val Pro Arg Gly Ala Arg Gln Pro Thr Leu Val Thr
        115                 120                 125

Ala Glu Arg Tyr Gly Tyr Val Trp Val Trp Tyr Gly Ser Pro Leu Pro
    130                 135                 140

Leu His Pro Leu Pro Glu Ile Ser Ala Ala Asp Val Asp Asn Gly Asp
145                 150                 155                 160

Phe Met His Leu His Phe Ala Phe Glu Thr Thr Thr Ala Val Leu Arg
                165                 170                 175

Ile Val Glu Asn Phe Tyr Asp Ala Gln His Ala Thr Pro Val His Ala
            180                 185                 190

Leu Pro Ile Ser Ala Phe Glu Leu Lys Leu Phe Asp Asp Trp Arg Gln
        195                 200                 205

Trp Pro Glu Val Glu Ser Leu Ala Leu Ala Gly Ala Trp Phe Gly Ala
    210                 215                 220

Gly Ile Asp Phe Thr Val Asp Arg Tyr Phe Gly Pro Leu Gly Met Leu
225                 230                 235                 240

Ser Arg Ala Leu Gly Leu Asn Met Ser Gln Met Asn Leu His Phe Asp
                245                 250                 255

Gly Tyr Pro Gly Gly Cys Val Met Thr Val Ala Leu Asp Gly Asp Val
            260                 265                 270
```

```
Lys Tyr Lys Leu Leu Gln Cys Val Thr Pro Val Ser Glu Gly Lys Asn
        275                 280                 285

Val Met His Met Leu Ile Ser Ile Lys Lys Val Gly Gly Ile Leu Leu
    290                 295                 300

Arg Ala Thr Asp Phe Val Leu Phe Gly Leu Gln Thr Arg Gln Ala Ala
305                 310                 315                 320

Gly Tyr Asp Val Lys Ile Trp Asn Gly Met Lys Pro Asp Gly Gly Gly
                325                 330                 335

Ala Tyr Ser Lys Tyr Asp Lys Leu Val Leu Lys Tyr Arg Ala Phe Tyr
            340                 345                 350

Arg Gly Trp Val Asp Arg Val Ala Ser Glu Arg
        355                 360
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28958 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
CGATCGCGTC GGCCTCGACA CCGTCGAAGA GGTCACGCTC GAAGCTCCCC TCGCTCTCCC     60

CTCTCAAGGC ACCATTCTCA TCCAGATCTC CGTCGGACCC ATGGACGAGG CGGGACGAAG    120

GTCGCTCTCC CTCCATGGCC GGACCGAGGA CGCTCCTCAG GACGCCCCTT GGACGCGCCA    180

CGCGAGCGGG TCGCTCGCTA AAGCTGCCCC CTCCCTCTCC TTCGATCTTC ACGAATGGGC    240

TCCTCCGGGG GGCACGCCGG TGGACACCCA AGGCTCTTAC GCAGGCCTCG AAAGCGGGGG    300

GCTCGCCTAT GGGCCTCAGT TCCAGGGACT TCGCTCCGTC TGGAAGCGCG GCGACGAGCT    360

CTTCGCCGAG GCCAAGCTCC CGGACGCAGG CGCCAAGGAT GCCGCTCGGT TCGCCCTCCA    420

CCCCGCCCTG TTCGACAGCG CCCTGCACGC GCTTGTCCTT GAAGACGAGC GGACGCCGGG    480

CGTCGCTCTG CCCTTCTCGT GGAGAGGAGT CTCGCTGCGC TCCGTCGGCG CCACCACCCT    540

GCGCGTGCGC TTCCATCGTC CGAATGGCAA GTCCTCCGTG TCGCTCCTCC TCGGCGACGC    600
```

CGCAGGCGAG CCCCTCGCCT CGGTCCAAGC GCTCGCCACG CGCATCACGT CCCAGGAGCA 660

GCTCCGCACC CAGGGAGCTT CCCTCCACGA TGCTCTCTTC CGGGTTGTCT GGAGAGATCT 720

GCCCAGCCCT ACGTCGCTCT CTGAGGCCCC GAAGGGTGTC CTCCTAGAGA CAGGGGGTCT 780

CGACCTCGCG CTGCAGGCGT CTCTCGCCCG CTACGACGGT CTCGCTGCCC TCCGGAGCGC 840

GCTCGACCAA GGCGCTTCGC CTCCGGGCCT CGTCGTCGTC CCCTTCATCG ATTCGCCCTC 900

TGGCGACCTC ATAGAGAGCG CTCACAACTC CACCGCGCGC GCCCTCGCCT TGCTGCAAGC 960

GTGGCTTGAC GACGAACGCC TCGCCTCCTC GCGCCTCGTC CTGCTCACCC GACAGGCCAT 1020

CGCAACCCAC CCCGACGAGG ACGTCCTCGA CCTCCCTCAC GCTCCTCTCT GGGGCCTTGT 1080

GCGCACCGCG CAAAGCGAAC ACCCGGAGCT CCCTCTCTTC CTCGTCGACC TGGACCTCGG 1140

TCAGGCCTCG GAGCGCGCCC TGCTCGGCGC GCTCGACACA GGAGAGCGTC AGCTCGCTCT 1200

CCGCCATGGA AAATGCCTCG TCCCGAGGTT GGTGAATGCA CGCTCGACAG AGGCGCTCAT 1260

CGCGCCGAAC GTATCCACGT GGAGCCTTCA TATCCCGACC AAAGGCACCT TCGACTCGCT 1320

CGCCCTCGTC GACGCTCCTC TAGCCCGTGC GCCCCTCGCA CAAGGCCAAG TCCGCGTCGC 1380

CGTGCACGCG GCAGGTCTCA ACTTCCGCGA TGTCCTCAAC ACCCTTGGCA TGCTTCCGGA 1440

CAACGCGGGG CCGCTCGGCG GCGAAGGCGC GGGCATTGTC ACCGAAGTCG GCCCAGGTGT 1500

TTCCCGATAC ACTGTAGGCG ACCGGGTGAT GGGCATCTTC CGCGGAGGCT TTGGCCCCAC 1560

GGTCGTCGCC GACGCCCGCA TGATCTGCCC CATCCCCGAT GCCTGGTCCT TCGTCCAAGC 1620

CGCCAGCGTC CCCGTCGTCT TTCTCACCGC CTACTATGGA CTCGTCGATG TCGGGCATCT 1680

CAAGCCCAAT CAACGTGTCC TCATCCATGC GGCCGCAGGC GGCGTCGGTA CTGCCGCCGT 1740

CCAGCTCGCG CGCCACCTCG GCGCCGAAGT CTTCGCCACC GCCAGTCCAG GGAAGTGGGA 1800

CGCTCTGCGC GCGCTCGGCT TCGACGATGC GCACCTCGCG TCCTCACGTG ACCTGGAATT 1860

CGAGCAGCAT TTCCTGCGCT CCACACGAGG GCGCGGCATG GATGTCGTCC TCAACGCCTT 1920

GGCGCGCGAG TTCGTCGACG CTTCGCTGCG TCTCCTGCCG AGCGGTGGAA GCTTTGTCGA 1980

GATGGGCAAG ACGGATATCC GCGAGCCCGA CGCCGTAGGC CTCGCCTACC CCGGCGTCGT 2040

TTACCGCGCC TTCGATCTCT TGGAGGCTGG ACCGGATCGA ATTCAAGAGA TGCTCGCAGA 2100

GCTGCTCGAC CTGTTCGAGC GCGGCGTGCT TCGTCCGCCG CCCATCACGT CCTGGGACAT 2160

CCGGCATGCC CCCCAGGCGT TCCGCGCGCT CGCTCAGGCG CGGCATATTG GAAAGTTCGT 2220

CCTCACCGTT CCCGTCCCAT CGATCCCCGA AGGCACCATC CTCGTCACGG GAGGCACCGG 2280
CACGCTCGGC GCGCTCATCG CGCGCCACCT CGTCGCCAAT CGCGGCGACA AGCACCTGCT 2340
CCTCACCTCG CGAAAGGGTG CGAGCGCTCC GGGGGCCGAG GCATTGCGGA GCGAGCTCGA 2400
AGCTCTGGGG GCTGCGGTCA CGCTCGCCCG GTGCGACGCG GCCGATCCAC GCGCGCTCCA 2460
AGCCCTCTTG GACAGCATCC CGAGCGCTCA CCCGCTCACG GCCGTCGTGC ACGCCGCCGG 2520
CGCCCTTGAC GATGGGCTGA TCAGCGACAT GAGCCCCGAG CGCATCGACC GCGTCTTTGC 2580
TCCCAAGCTC GACGCCGCTT GGCACTTGCA TCAGCTCACC CAGGACAAGG CCGCTCGGGG 2640
CTTCGTCCTC TTCTCGTCCG CCTCCGGCGT CCTCGGCGGT ATGGGTCAAT CCAACTACGC 2700
GGGGGGCAAT GCGTTCCTTG ACGCGCTCGC GCATCACCGA CGCGTCCATG GGCTCCCAGG 2760
CTCCTCGCTC GCATGGGGCC ATTGGGCCGA GCGCAGCGGA ATGACCCGAC AACCTCAGCG 2820
GCGTCGATAC CGCTCGCATG AGGCGCGCGG TCTCCGATCC ATCGCCTCGG ACGAGGGTCT 2880
CGCCCTCTTC GATATGGCGC TCGGGCGCCC GGAGCCCGCG CTGGTCCCCG CCCGCTTCGA 2940
CATGAACGCG CTCGGCGCGA AGGCCGACGG GCTACCCTCG ATGTTCCAGG GTCTCGTCCG 3000
CGCTCGCGTC GCGCGCAAGG TCGCCAGCAA TAATGCCCTG GCCGCGTCGC TCACCCAGCG 3060
CCTCGCCTCC CTCCCGCCCA CCGACCGCGA GCGCATGCTG CTCGATCTCG TCCGCGCCGA 3120
AGCCGCCATC GTCCTCGGCC TCGCCTCGTT CGAATCGCTC GATCCCCGTC GCCCTCTTCA 3180
AGAGCTCGGT CTCGATTCCC TCATGGCCAT CGAGCTCCGA AATCGACTCG CCGCCGCCAC 3240
AGGCTTGCGA CTCCAAGCCA CCCTCCTCTT CGACCACCCG ACGCCCGCCG CGCTCGCGAC 3300
CCTGCTGCTC GGGAAGCTCC TCCAGCATGA AGCTGCCGAT CCTCGCCCCT TGGCCGCAGA 3360
GCTCGACAGG CTAGAGGCCA CTCTCTCCGC GATAGCCGTG GACGCTCAAG CACGCCCGAA 3420
GATCATATTA CGCCTGCAAT CCTGGTTGTC GAAGTGGAGC GACGCTCAGG CTGCCGACGC 3480
TGGACCGATT CTCGGCAAGG ATTTCAAGTC TGCTACGAAG GAAGAGCTCT TCGCTGCTTG 3540
TGACGAAGCG TTCGGAGGCC TGGGTAAATG AATAACGACG AGAAGCTTGT CTCCTACCTA 3600
CAGCAGGCGA TGAATGAGCT TCAGCGTGCT CATCAGCCCC TCCGCGCGGT CGAAGAGAAG 3660
GAGCACGAGC CCATCGCCAT CGTGGCGATG AGCTGCCGCT TCCCGGGCGA CGTGCGCACG 3720
CCCGAGGATC TCTGGAAGCT CTTGCTCGAT GGGAAAGATG CTATCTCCGA CCTTCCCCCA 3780
AACCGTGGTT GGAAGCTCGA CGCGCTCGAC GTCCACGGTC GCTCCCCAGT CCGAGAGGGA 3840
GGCTTCTTCT ACGACGCAGA CGCCTTCGAT CCGGCCTTCT TCGGGATCAG CCCACGCGAG 3900

GCGCTCGCCA TCGATCCCCA GCAGCGGCTC CTCCTCGAGA TCTCATGGGA AGCCTTCGAG 3960

CGTGCGGGCA TCGACCCTGC CTCGCTCCAA GGGAGCCAAA GCGGCGTCTT CGTCGGCGTG 4020

ATACACAACG ACTACGACGC ATTGCTGGAG AACGCAGCTG GCGAACACAA AGGATTCGTT 4080

TCCACCGGCA GCACAGCGAG CGTCGCCTCC GGCCGGATCG CGTATACATT CGGCTTTCAA 4140

GGGCCCGCCA TCAGCGTGGA CACGGCGTGC AGCTCCTCGC TCGTCGCGGT TCACCTCGCC 4200

TGCCAGGCCC TGCGCCGTGG CGAATGCTCC CTGGCGCTCG CCGGCGGCGT GACCGTCATG 4260

GCCACGCCAG CAGTCTTCGT CGCGTTCGAT TCCGAGAGCG CGGGCGCCCC CGATGGTCGC 4320

TGCAAGTCGT TCTCGGTGGA GGCCAACGGT TCGGGCTGGG CCGAGGGCGC CGGGATGCTC 4380

CTGCTCGAGC GCCTCTCCGA TGCCGTCCAA AACGGTCATC CCGTCCTCGC CGTCCTTCGA 4440

GGCTCCGCCG TCAACCAGGA CGGCCGGAGC CAAGGCCTCA CCGCGCCCAA TGGCCCTGCC 4500

CAAGAGCGCG TCATCCGGCA AGCGCTCGAC AGCGCGCGGC TCACTCCAAA GGACGTCGAC 4560

GTCGTCGAGG CTCACGGCAC GGGAACCACC CTCGGAGACC CCATCGAGGC ACAGGCCATT 4620

CTTGCCACCT ATGGCGAGGC CCATTCCCAA GACAGACCCC TCTGGCTTGG AAGTCTCAAG 4680

TCCAACCTGG GACATGCTCA GGCCGCGGCC GGCGTGGGAA GCGTCATCAA GATGGTGCTC 4740

GCGTTGCAGC AAGGCCTCTT GCCCAAGACC CTCCATGCCC AGAATCCCTC CCCCCACATC 4800

GACTGGTCTC CGGGCACGGT AAAGCTCCTG AACGAGCCCG TCGTCTGGAC GACCAACGGG 4860

CATCCTCGCC ACGCCGGCGT CTCCGCCTTC GGCATCTCCG GCACCAACGC CCACGTCATC 4920

CTCGAAGAGG CCCCCGCCAT CGCCCGGGTC GAGCCCGCAG CGTCACAGCC CGCGTCCGAG 4980

CCGCTTCCCG CAGCGTGGCC CGTGCTCCTG TCGGCCAAGA GCGAGGCGGC CGTGCGCGCC 5040

CAGGCAAAGC GGCTCCGCGA CCACCTCCTC GCCAAAAGCG AGCTCGCCCT CGCCGATGTG 5100

GCCTATTCGC TCGCGACCAC GCGCGCCCAC TTCGAGCAGC GCGCCGCTCT CCTCGTCAAA 5160

GGCCGCGACG AGCTCCTCTC CGCCCTCGAT GCGCTGGCCC AAGGACATTC CGCCGCCGTG 5220

CTCGGACGAA GCGGGGCCCC AGGAAAGCTC GCCGTCCTCT TCACGGGGCA AGGAAGCCAG 5280

CGGCCCACCA TGGGCCGCGG CCTCTACGAC GTTTTCCCCG TCTTCCGGGA CGCCCTCGAC 5340

ACCGTCGGCG CCCACCTCGA CCGCGAGCTC GACCGCCCCC TGCGCGACGT CCTCTTCGCT 5400

CCCGACGGCT CCGAGCAGGC CGCGCGCCTC GAGCAAACCG CCTTCACCCA GCCGGCCCTG 5460

TTTGCCCTCG AAGTCGCCCT CTTTCAGCTT CTACAATCCT TCGGTCTGAA GCCCGCTCTC 5520

CTCCTCGGAC ACTCCATTGG CGAGCTCGTC GCCGCCCACG TCGCCGGCGT CCTTTCTCTC 5580

CAGGACGGCT GCACCCTCGT CGCCGCCCGC GCAAAGCTCA TGCAAGCGCT CCCACAAGGC 5640

GGCGCCATGG TCACCCTCCG AGCCTCCGAG GAGGAAGTCC GCGACCTTCT CCAGCCCTAC 5700

GAAGGCCGAG CTAGCCTCGC CGCCCTCAAT GGGCCTCTCT CCACCGTCGT CGCTGGCGAT 5760

GAAGACGCGG TGGTGGAGAT CGCCCGCCAG GCCGAAGCCC TCGGACGAAA GACCACACGC 5820

CTGCGCGTCA GCCACGCCTT CCATTCCCCG CACATGGACG GAATGCTCGA CGACTTCCGC 5880

CGCGTCGCCC AGAGCCTCAC CTACCATCCC GCACGCATCC CCATCATCTC CAACGTCACC 5940

GGCGCGCGCG CCACGGACCA CGAGCTCGCC TCGCCCGACT ACTGGGTCCG CCACGTTCGC 6000

CACACCGTCC GCTTCCTCGA CGGCGTACGT GCCCTTCACG CCGAAGGGGC ACGTGTCTTT 6060

CTCGAGCTCG GGCCTCACGC TGTCCTCTCC GCCCTTGCGC AAGACGCCCT CGGACAGGAC 6120

GAAGGCACGT CGCCATGCGC CTTCCTTCCC ACCCTCCGCA AGGGACGCGA CGACGCCGAG 6180

GCGTTCACCG CCGCGCTCGG CGCTCTCCAC TCCGCAGGCA TCACACCCGA CTGGAGCGCT 6240

TTCTTCGCCC CCTTCGCTCC ACGCAAGGTC TCCCTCCCCA CCTATGCCTT CCAGCGCGAG 6300

CGCTTCTGGC CCGACGCCTC CAAGGCACCC GGCGCCGACG TCAGCCACCT TGCTCCGCTC 6360

GAGGGGGGGC TCTGGCAAGC CATCGAGCGC GGGGACCTCG ATGCGCTCAG CGGTCAGCTC 6420

CACGTGGACG GCGACGAGCG GCGCGCCGCG CTCGCCCTGC TCCTTCCCAC CCTCTCGAGC 6480

TTTCGCCACG AGCGGCAAGA GCAGAGCACG GTCGACGCCT GGCGCTACCG TATCACCTGG 6540

AAGCCTCTGA CCACCGCCGA AACACCCGCC GACCTCGCCG GCACCTGGCT CGTCGTCGTG 6600

CCGGCCGCTC TGGACGACGA CGCGCTCCCC TCCGCGCTCA CCGAGGCGCT CACCCGGCGC 6660

GGCGCGCGCG TCCTCGCCTT GCGCCTGAGC CAGGCCCACC TGGACCGCGA GGCTCTCGCC 6720

GAGCATCTGC GCCAGGCTTG CGCCGAGACC GCCCCGATTC GCGGCGTGCT CTCGCTCCTC 6780

GCCCTCGACG AGCGCCCCCT CGCAGACCGT CCTGCCCTGC CCGCCGGACT CGCCCTCTCG 6840

CTTTCTCTCG CTCAAGCCCT CGGCGACCTC GACCTCGAGG CGCCCTTGTG GTTCTTCACG 6900

CGCGGCGCCG TCTCCATTGG ACACTCTGAC CCCCTCGCCC ATCCCGCCCA GGCCATGACC 6960

TGGGGCTTGG GCCGCGTCAT CGGCCTCGAG CACCCCGACC GGTGGGGAGG TCTCGTCGAC 7020

GTCTGCGCTG GGGTCGACGA GAGCGCCGTG GGCCGCTTGC TGCCGGCCCT CGCCGAGCGC 7080

CACGACGAAG ACCAGCTCGC TCTCCGCCCG GCCGGACTCT ACGCTCGCCG CATCGTCCGC 7140

GCCCCGCTCG GCGATGCGCC TCCCGCGCGC GACTTCACGC CCGGAGGCAC CATTCTCATC 7200

ACCGGCGGCA CCGGCGCCAT TGGCGCTCAC GTCGCCCGAT GGCTCGCTCG AAGAGGCGCT 7260

CAGCACCTCG TCCTCATCAG CCGCCGAGGC GCCGAGGCCC CTGGCGCCTC GGAGCTCCAC 7320

GACGAGCTCT CGGCCCTCGG CGCGCGCACC ACCCTCGCCG CGTGCGATGT CGCCGACCGG 7380

AATGCTGTCG CCACGCTTCT TGAGCAGCTC GACGCCGAAG GGTCGCAGGT CCGCGCCGTG 7440

TTCCACGCGA GCGGCATCGA ACACCACGCT CCGCTCGACG CCACCTCTTT CAGGGATCTC 7500

GCCGAGGTTG TCTCCGGCAA GGTCGAAGGT GCAAAGCACC TCCACGACCT GCTCGGCTCT 7560

CGACCCCTCG ACGCCTTTGT TCTCTTTTCG TCCGGCGCGG CCGTCTGGGG CGGCGGACAG 7620

CAAGGCGGCT ACGCGGCCGC AAACGCCTTC CTCGACGCCC TTGCCGAGCA TCGGCGCAGC 7680

GCTGGATTGA CAGCGACGTC GGTGGCCTGG GGCGCGTGGG GCGGCGGCGG CATGGCCACC 7740

GATCAGGCGG CAGCCCACCT CCAACAGCGC GGTCTGTCGC GGATGGCCCC CTCGCTTGCC 7800

CTGGCGGCGC TCGCGCTGGC TCTGGAGCAC GACGAGACCA CCGTCACCGT CGCCGACATC 7860

GACTGGGCGC GCTTTGCGCC TTCGTTCAGC GCCGCTCGCC CCCGCCCGCT CCTGCGCGAT 7920

TTGCCCGAGG CGCAGCGCGC TCTCGAGACC AGCGAAGGCG CGTCCTCCGA GCATGGCCCG 7980

GCCCCCGACC TCCTCGACAA GCTCCGGAGC CGCTCGGAGA GCGAGCAGCT TCGTCTGCTC 8040

GTCTCGCTGG TGCGCCACGA GACGGCCCTC GTCCTCGGCC ACGAAGGCGC CTCCCATGTC 8100

GACCCCGACA AGGGCTTCCT CGATCTCGGT CTCGATTCGC TCATGGCCGT CGAGCTTCGC 8160

CGGCGCTTGC AACAGGCCAC CGGCATCAAG CTCCCGGCCA CCCTCGCCTT CGACCATCCC 8220

TCTCCTCATC GAGTCGCGCT CTTCTTGCGC GACTCGCTCG CCCACGCCCT CGGCACGAGG 8280

CTCTCCGTCG AGCCCGACGC CGCCGCGCTC CCGGCGCTTC GCGCCGCGAG CGACGAGCCC 8340

ATCGCCATCG TCGGCATGGC CCTCCGCCTG CCGGGCGGCG TCGGCGATGT CGACGCTCTT 8400

TGGGAGTTCC TGGCCCAGGG ACGCGACGGC GTCGAGCCCA TTCCAAAGGC CCGATGGGAT 8460

GCCGCTGCGC TCTACGACCC CGACCCCGAC GCCAAGACCA AGAGCTACGT CCGGCATGCC 8520

GCCATGCTCG ACCAGGTCGA CCTCTTCGAC CCTGCCTTCT TTGGCATCAG CCCCCGGGAG 8580

GCCAAACACC TCGACCCCCA GCACCGCCTG CTCCTCGAAT CTGCCTGGCA GGCCCTCGAA 8640

GACGCCGGCA TCGTCCCCCC CACCCTCAAG GATTCCCCCA CCGGCGTCTT CGTCGGCATC 8700

GGCGCCAGCG AATACGCATT GCGAGAGGCG AGCACCGAAG ATTCCGACGC TTATGCCCTC 8760

CAAGGCACCG CCGGGTCCTT TGCCGCGGGG CGCTTGGCCT ACACGCTCGG CCTGCAAGGG 8820

CCCGCGCTCT CGGTCGACAC CGCCTGCTCC TCCTCGCTCG TCGCCCTCCA CCTCGCCTGC 8880
CAAGCCCTCC GACAGGGCGA GTGCAACCTC GCCCTCGCCG CGGGCGTCTC CGTCATGGCC 8940
TCCCCCGAGG GCTTCGTCCT CCTTTCCCGC CTGCGCGCCT TGGCGCCCGA CGGCCGCTCC 9000
AAGACCTTCT CGGCCAACGC CGACGGCTAC GGACGCGGAG AAGGCGTCAT CGTCCTTGCC 9060
CTCGAGCGGC TCGGTGACGC CCTCGCCCGA GGACACCGCG TCCTCGCCCT CGTCCGCGGC 9120
ACCGCCATCA ACCACGACGG CGCGTCGAGC GGTATCACCG CCCCCAACGG CACCTCCCAG 9180
CAGAAGGTCC TCCGCGCCGC GCTCCACGAC GCCCGCATCA CCCCCGCCGA CGTCGACGTC 9240
GTCGAGTGCC ATGGCACCGG CACCTCCTTG GGAGACCCCA TCGAGGTGCA AGCCCTGGCC 9300
GCCGTCTACG CCGACGGCAG ACCCGCTGAA AAGCCTCTCC TTCTCGGCGC GCTCAAGACC 9360
AACATCGGCC ATCTCGAGGC CGCCTCCGGC CTCGCGGGCG TCGCCAAGAT CGTCGCCTCC 9420
CTCCGCCATG ACGCCCTGCC CCCCACCCTC CACACGGGCC CGCGCAATCC CTTGATTGAT 9480
TGGGATACAC TCGCCATCGA CGTCGTTGAT ACCCCGAGGT CTTGGGCCCG CCACGAAGAT 9540
AGCAGTCCCC GCCGCGCCGG CGTCTCCGCC TTCGGACTCT CCGGCACCAA CGCCCACGTC 9600
ATCCTCGAGG AGGCTCCCGC CGCCCTGTCG GGCGAGCCCG CCACCTCACA GACGGCGTCG 9660
CGACCGCTCC CCGCGGCGTG TGCCGTGCTC CTGTCGGCCA GGAGCGAGGC CGCCGTCCGC 9720
GCCCAGGCGA AGCGGCTCCG CGACCACCTC CTCGCCCACG ACGACCTCGC CCTTATCGAT 9780
GTGGCCTATT CGCAGGCCAC CACCCGCGCC CACTTCGAGC ACCGCGCCGC TCTCCTGGCC 9840
CGCGACCGCG ACGAGCTCCT CTCCGCGCTC GACTCGCTCG CCCAGGACAA GCCCGCCCCG 9900
AGCACCGTTC TCGGCCGGAG CGGAAGCCAC GGCAAGGTCG TCTTCGTCTT TCCTGGGCAA 9960
GGCTCGCAGT GGGAAGGGAT GGCCCTCTCC CTGCTCGACT CCTCGCCGGT CTTCCGCGCT 10020
CAGCTCGAAG CATGCGAGCG CGCGCTCGCT CCTCACGTCG AGTGGAGCCT GCTCGCCGTC 10080
CTGCGCCGCG ACGAGGGCGC CCCCTCCCTC GACCGCGTCG ACGTCGTACA GCCCGCCCTC 10140
TTTGCCGTCA TGGTCTCCCT GGCCGCCCTC TGGCGCTCGC TCGGCGTCGA GCCCGCCGCC 10200
GTCGTCGGCC ACAGCCAGGG CGAGATCGCC GCCGCCTTCG TCGCAGGCGC TCTCTCCCTC 10260
GAGGACGCGG CGCGCATCGC CGCCCTGCGC AGGAAAGCGC TCACCACCGT CGGCGGCAAC 10320
GGCGGCATGG CCGCCGTCGA GCTCGGCGCC TCCGACCTCC AGACCTACCT CGCTCCCTGG 10380
GGCGACAGGC TCTCCACCGC CGCCGTCAAC AGCCCCAGGG CTACCCTCGT ATCCGGCGAG 10440
CCCGCCGCCG TCGACGCGCT GCTCGACGTC CTCACCGCCA CCAAGGTGTT CGCCCGCAAG 10500

ATCCGCGTCG ACTACGCCTC CCACTCCGCC CAGATGGACG CCGTCCAAGA CGAGCTCGCC 10560

GCAGGTCTAG CCAACATCGC TCCTCGGACG TGCGAGCTCC CTCTTTATTC GACCGTCACC 10620

GGCACCAGGC TCGACGGCTC CGAGCTCGAC GGCGCGTACT GGTATCGAAA CCTCCGGCAA 10680

ACCGTCCTGT TCTCGAGCGC GACCGAGCGG CTCCTCGACG ATGGGCATCG CTTCTCCGTC 10740

GAGGTCAGCC CCCATCCCGT GCTCACGCTC GCCCTCCGCG AGACCTGCGA GCGCTCACCG 10800

CTCGATCCCG TCGTCGTCGG CTCCATTCGA CGAGAAGAAG GCCACCTCGC CCGCCTGCTC 10860

CTCTCCTGGG CGGAGCTCTC TACCCGAGGC CTCGCGCTCG ACTGGAAGGA CTTCTTCGCG 10920

CCCTACGCTC CCCGCAAGGT CTCCCTCCCC ACCTACCCCT TCCAGCGAGA GCGGTTCTGG 10980

CTCGACGTCT CCACGGACGA ACGCTTCCGA CGTCGCCTCC GCAGGCCTGA CCTCGGCCGA 11040

CCAATCCCGC TGCTCGGCGC CGCCGTCGCC TTCGCCGACC GCGGTGGCTT TCTCTTTACA 11100

GGGCGGCTCT CCCTCGCAGA GCACCCGTGG CTCGAAGGCC ATGCCGTCTT CGGCACACCC 11160

ATCCTACCGG GCACCGGCTT TCTCGAGCTC GCCCTGCACG TCGCCCACCG CGTCGGCCTC 11220

GACACCGTCG AAGAGCTCAC GCTCGAGGCC CCTCTCGCTC TCCCATCGCA GGACACCGTC 11280

CTCCTCCAGA TCTCCGTCGG GCCCGTGGAC GACGCAGGAC GAAGGGCGCT CTCTTTCCAT 11340

AGCCGACAAG AGGACGCGCT TCAGGATGGC CCCTGGACTC GCCACGCCAG CGGCTCTCTC 11400

TCGCCGGCGA CCCCATCCCT CTCCGCCGAT CTCCACGAGT GGCCTCCCTC GAGTGCCATC 11460

CCGGTGGACC TCGAAGGCCT CTACGCAACC CTCGCCAACC TCGGGCTTGC CTACGGCCCC 11520

GAGTTCCAGG GCCTCCGCTC CGTCTACAAG CGCGGCGACG AGCTCTTTGC CGAAGCCAAG 11580

CTCCCGGAAG CGGCCGAAAA GGATGCCGCC CGGTTTGCCC TCCACCCTGC GCTGCTCGAC 11640

AGCGCCCTGC ATGCACTGGC CTTTGAGGAC GAGCAGAGAG GGACGGTCGC TCTGCCCTTC 11700

TCGTGGAGCG GAGTCTCGCT GCGCTCCGTC GGTGCCACCA CCTTGCGCGT GCGCTTCCAC 11760

CGTCCCAAGG GTGAATCCTC CGTCTCGATC GTCCTGGCCG ACGCCGCAGG TGACCCTCTT 11820

GCCTCGGTGC AAGCGCTCGC CATGCGGACG ACGTCCGCCG CGCAGCTCCG CACCCCGGCA 11880

GCTTCCCACC ATGATGCGCT CTTCCGCGTC GACTGGAGCG AGCTCCAAAG CCCCACTTCA 11940

CCGCCTGCCG CCCCGAGCGG CGTCCTTCTC GGCACAGGCG GCCACGATCT CGCGCTCGAC 12000

GCCCCGCTCG CCCGCTACGC CGACCTCGCT GCCCTCCGAA GCGCCCTCGA CCAGGGCGCT 12060

TCGCCTCCCG GCCTCGTCGT CGCCCCCTTC ATCGATCGAC CGGCAGGCGA CCTCGTCCCG 12120

AGCGCCCACG AGGCCACCGC GCTCGCACTC GCCCTCTTGC AAGCCTGGCT CGCCGACGAA 12180

CGCCTCGCCT CGTCGCGCCT CGTCCTCGTC ACCCGACGCG CCGTCGCCAC CCACACCGAA 12240

GACGACGTCA AGGACCTCGC TCACGCGCCG CTCTGGGGGC TCGCGCGCTC CGCGCAAAGT 12300

GAGCACCCAG ACCTCCCGCT CTTCCTCGTC GACATCGACC TCAGCGAGGC CTCCCAGCAG 12360

GCCCTGCTAG GCGCGCTCGA CACAGGAGAA CGCCAGCTCG CCCTCCGCAA CGGGAAACCC 12420

CTCATCCCGA GGTTGGCGCA ACCACGCTCG ACGGACGCGC TCATCCCGCC GCAAGCACCC 12480

ACGTGGCGCC TCCATATTCC GACCAAAGGC ACCTTCGACG CGCTCGCCCT CGTCGACGCC 12540

CCCGAGGCCC AGGCGCCCCT CGCACACGGC CAAGTCCGCA TCGCCGTGCA CGCGGCAGGG 12600

CTCAACTTCC GCGATGTCGT CGACACCCTT GGCATGTATC CGGGCGACGC GCCGCCGCTC 12660

GGAGGCGAAG GCGCGGGCAT CGTTACTGAA GTCGGTCCAG GTGTCTCCCG ATACACCGTA 12720

GGCGACCGGG TGATGGGGGT CTTCGGCGCA GCCTTTGGTC CCACGGCCAT CGCCGACGCC 12780

CGCATGATCT GCCCCATCCC CCACGCCTGG TCCTTCGCCC AAGCCGCCAG CGTCCCCATC 12840

ATCTATCTCA CCGCCTACTA TGGACTCGTC GATCTCGGGC ATCTGAAACC CAATCAACGT 12900

GTCCTCATCC ATGCGGCCGC CGGCGGCGTC GGGACGGCCG CCGTTCAGCT CGCACGCCAC 12960

CTCGGCGCCG AGGTCTTTGC CACCGCCAGT CCAGGGAAGT GGAGCGCTCT CCGCGCGCTC 13020

GGCTTCGACG ATGCGCACCT CGCGTCCTCA CGTGACCTGG GCTTCGAGCA GCACTTCCTG 13080

CGCTCCACGC ATGGGCGCGG CATGGATGTC GTCCTCGACT GTCTGGCACG CGAGTTCGTC 13140

GACGCCTCGC TGCGCCTCAT GCCGAGCGGT GGACGCTTCA TCGAGATGGG AAAGACGGAC 13200

ATCCGTGAGC CCGACGCGAT CGGCCTCGCC TACCCTGGCG TCGTTTACCG CGCCTTCGAC 13260

GTCACAGAGG CCGGACCGGA TCGAATTGGG CAGATGCTCG CAGAGCTGCT CAGCCTCTTC 13320

GAGCGCGGTG TGCTTCGTCT GCCACCCATC ACATCCTGGG ACATCCGTCA TGCCCCCCAG 13380

GCCTTCCGCG CGCTCGCCCA GGCGCGGCAT GTTGGGAAGT TCGTCCTCAC CATTCCCCGT 13440

CCGATCGATC CCGAGGGGAC CGTCCTCATC ACGGGAGGCA CCGGGACGCT AGGAGTCCTG 13500

GTCGCACGCC ACCTCGTCGC GAAACACAGC GCCAAACACC TGCTCCTCAC CTCGAGGAAG 13560

GGCGCGCGTG CTCCGGGCGC GGAGGCTCTG CGAAGCGAGC TCGAAGCGCT GGGGGCCTCG 13620

GTCACCCTCG TCGCGTGCGA CGTGGCCGAC CCACGCGCCC TCCGGACCCT CCTGGACAGC 13680

ATCCCGAGGG ATCATCCGAT CACGGCCGTC GTGCACGCCG CCGGCGCCCT CGACGACGGG 13740

CCGCTCGGTA GCATGAGCGC CGAGCGCATC GCTCGCGTCT TTGACCCCAA GCTCGATGCC 13800

GCTTGGTACT TGCATGAGCT CACCCAGGAC GAGCCGGTCG CGGCCTTCGT CCTCTTCTCG 13860

GCCGCCTCCG GCGTCCTTGG TGGTCCAGGT CAGTCGAACT ACGCCGCTGC CAATGCCTTC 13920

CTCGATGCGC TCGCACATCA CCGGCGCGCC CAAGGACTCC CAGCCGCTTC GCTCGCCTGG 13980

GGCTACTGGG CCGAGCGCAG TGGGATGACC CGGCACCTCA GCGCCGCCGA CGCCGCTCGC 14040

ATGAGGCGCG CCGGCGTCCG GCCCCTCGAC ACTGACGAGG CGCTCTCCCT CTTCGATGTG 14100

GCTCTCTTGC GACCCGAGCC CGCTCTGGTC CCCGCCCCCT TCGACTACAA CGTGCTCAGC 14160

ACGAGTGCCG ACGGCGTGCC CCCGCTGTTC CAGCGTCTCG TCCGCGCTCG CATCGCGCGC 14220

AAGGCCGCCA GCAATACTGC CCTCGCCTCG TCGCTTGCAG AGCACCTCTC CTCCCTCCCG 14280

CCCGCCGAAC GCGAGCGCGT CCTCCTCGAT CTCGTCCGCA CCGAAGCCGC CTCCGTCCTC 14340

GGCCTCGCCT CGTTCGAATC GCTCGATCCC CATCGCCCTC TACAAGAGCT CGGCCTCGAT 14400

TCCCTCATGG CCCTCGAGCT CCGAAATCGA CTCGCCGCCG CCGCCGGGCT GCGGCTCCAG 14460

GCTACTCTCC TCTTCGACTA TCCAACCCCG ACTGCGCTCT CACGCTTTTT CACGACGCAT 14520

CTCTTCGGGG GAACCACCCA CCGCCCCGGC GTACCGCTCA CCCCGGGGGG GAGCGAAGAC 14580

CCTATCGCCA TCGTGGCGAT GAGCTGCCGC TTCCCGGGCG ACGTGCGCAC GCCCGAGGAT 14640

CTCTGGAAGC TCTTGCTCGA CGGACAAGAT GCCATCTCCG GCTTTCCCCA AAATCGCGGC 14700

TGGAGTCTCG ATGCGCTCGA CGCCCCCGGT CGCTTCCCAG TCCGGGAGGG GGGCTTCGTC 14760

TACGACGCAG ACGCCTTCGA TCCGGCCTTC TTCGGGATCA GTCCACGTGA AGCGCTCGCC 14820

GTTGATCCCC AACAGCGCAT TTTGCTCGAG ATCACATGGG AAGCCTTCGA GCGTGCAGGC 14880

ATCGACCCGG CCTCCCTCCA AGGAAGCCAA AGCGGGGTCT TCGTTGGCGT ATGGCAGAGC 14940

GACTACCAAT GCATCGCTGG TGAACGCGAC TGGCGAATAC AAGGACTCGT TGCCACCGGT 15000

AGCGCAGCGC GTCCGTCCGG CCGAATCGCA TACACGTTCG GACTTCAAGG GCCCGCCATC 15060

AGCGTGGAGA CGGCGTGCAG CTTCCTCGTC GCGGTTCACC TCGCCTGCCA GGCCCCCCCC 15120

CACGGCGAAT ACTCCCTGGC GCTCGCTGGC GGCGTGACCA TCATGGCCAC GCCAGCCATA 15180

TTCATCGCGT TCGACTCCGA GAGCGCGGGT GCCCCCGACG GTCGCTGCAA GGCCTTCTCG 15240

CCGGAAGCCG ACGGTTCGGG CTGGGCCGAA GGCGCCGGGA TGCTCCTGCT CGAGCGCCTC 15300

TCCGATGCCG TCCAAAACGG TCATCCCGTC CTCGCCGTCC TTCGAGGCTC CGCCGTCAAC 15360

CAGGACGGCC GGAGCCAAGG CCTCACCGCG CCCAATGGCC CTGCCCAGGA GCGCGTCATC 15420

CGGCAAGCGC TCGACAGCGC GCGGCTCACT CCAAAGGACG TCGACGTCGT CGAGGCTCAC 15480

GGCACGGGAA CCACCCTCGG AGACCCCATC GAGGCACAGG CCGTTTTTGC CACCTATGGC 15540

GAGGCCCATT CCCAAGACAG ACCCCTCTGG CTTGGAAGCC TCAAGTCCAA CCTGGGACAT 15600

ACTCAGGCCG CGGCCGGCGT CGGCGGCATC ATCAAGATGG TGCTCGCGTT GCAGCACGGT 15660

CTCTTGCCCA AGACCCTCCA TGCCCAGAAT CCCTCCCCCC ACATCGACTG GTCTCCAGGC 15720

ATCGTAAAGC TCCTGAACGA GGCCGTCGCC TGGACGACCA GCGGACATCC TCGCCGCGCC 15780

GGTGTTTCCT CGTTCGGCGT CTCCGGCACC AACGCCCATG TCATCCTCGA AGAGGCTCCC 15840

GCCGCCACGC GGGCCGAGTC AGGCGCTTCA CAGCCTGCAT CGCAGCCGCT CCCCGCGGCG 15900

TGGCCCGTCG TCCTGTCGGC CAGGAGCGAG GCCGCCGTCC GCGCCCAGGC TCAAAGGCTC 15960

CGCGAGCACC TGCTCGCCCA AGGCGACCTC ACCCTCGCCG ATGTGGCCTA TTCGCTGGCC 16020

ACCACCCGCG CCCACTTCGA GCACCGCGCC GCTCTCGTAG CCCACGACCG CGACGAGCTC 16080

CTCTCCGCGC TCGACTCGCT CGCCCAGGAC AAGCCCGCAC CGAGCACCGT CCTCGGACGG 16140

AGCGGAAGCC ACGGCAAGGT CGTCTTCGTC TTTCCTGGGC AAGGCTCGCA GTGGGAAGGG 16200

ATGGCCCTCT CCCTGCTCGA CTCCTCGCCC GTCTTCCGCA CACAGCTCGA AGCATGCGAG 16260

CGCGCGCTCC GTCCTCACGT CGAGTGGAGC CTGCTCGCCG TCCTGCGCCG CGACGAGGGC 16320

GCCCCCTCCC TCGACCGCGT CGACGTCGTG CAGCCCGCCC TCTTTGCCGT CATGGTCTCC 16380

CTGGCCGCCC TCTGGCGCTC GCTCGGCGTC GAGCCCGCCG CCGTCGTCGG CCACAGCCAG 16440

GGCGAGATAG CCGCCGCCTT CGTCGCAGGC GCTCTCTCCC TCGAGGACGC GGCCCGCATC 16500

GCCGCCCTGC GCAGCAAAGC GTCACCACCG TCGCCGGCAA CGGGCATGGC CGCCGTCGAG 16560

CTCGGCGCCT CCGACCTCCA GACCTACCTC GCTCCCTGGG GCGACAGGCT CTCCATCGCC 16620

GCCGTCAACA GCCCCAGGGC CACGCTCGTA TCCGGCGAGC CCGCCGCCGT CGACGCGCTG 16680

ATCGACTCGC TCACCGCAGC GCAGGTCTTC GCCCGAAGAG TCCGCGTCGA CTACGCCTCC 16740

CACTCAGCCC AGATGGACGC CGTCCAAGAC GAGCTCGCCG CAGGTCTAGC CAACATCGCT 16800

CCTCGGACGT GCGAGCTCCC TCTTTATTCG ACCGTCACCG GCACCAGGCT CGACGGCTCC 16860

GAGCTCGACG GCGCGTACTG GTATCGAAAC CTCCGGCAAA CCGTCCTGTT CTCGAGCGCG 16920

ACCGAGCGGC TCCTCGACGA TGGGCATCGC TTCTTCGTCG AGGTCAGCCC TCATCCCGTG 16980

CTCACGCTCG CCCTCCGCGA GACCTGCGAG CGCTCACCGC TCGATCCCGT CGTCGTCGGC 17040

TCCATTCGAC GCGACGAAGG CCACCTCCCC CGTCTCCTTG CTCTCTTGGG CCGAGCTCTA 17100

TGGCCGGGCC TCACGCCCGA GTGGAAGGCC TTCTTCGCGC CCTTCGCTCC CCGCAAGGTC 17160

TCACTCCCCA CCTACGCCTT CCAGCGCGAG CGTTTCTGGC TCGACGCCCC CAACGCACAC 17220

CCCGAAGGCG TCGCTCCCGC TGCGCCGATC GATGGGCGGT TTTGGCAAGC CATCGAACGC 17280

GGGGACCTCG ACGCGCTCAG CGGCCAGCTC CACGCGGACG GCGACGAGCA GCGCGCCGCC 17340

CTCGCCCTGC TCCTTCCCAC CCTCTCGAGC TTTCACCACC AGCGCCAAGA GCAGAGCACG 17400

GTCGACACCT GGCGCTACCG CATCACGTGG AGGCCTCTGA CCACCGCCGC CACGCCCGCC 17460

GACCTCGCCG GCACCTGGCT CCTCGTCGTG CCGTCCGCGC TCGGCGACGA CGCGCTCCCT 17520

GCCACGCTCA CCGATGCGCT TACCCGGCGC GGCGCGCGTG TCCTCGCGCT GCGCCTGAGC 17580

CAGGTTCACA TAGGCCGCGC GGCTCTCACC GAGCACCTGC GCGAGGCTGT TGCCGAGACT 17640

GCCCCGATTC GCGGCGTGCT CTCCCTCCTC GCCCTCGACG AGCGCCCCCT CGCGGACCAT 17700

GCCGCCCTGC CCGCGGGCCT TGCCCTCTCG CTCGCCCTCG TCCAAGCCCT CGGCGACCTC 17760

GCCCTCGAGG CTCCCTTGTG GCTCTTCACG CGCGGCGCCG TCTCGATTGG ACACTCCGAC 17820

CCACTCGCCC ATCCCACCCA GGCCATGATC TGGGGCTTGG GCCGCGTCGT CGGCCTCGAG 17880

CACCCCGAGC GGTGGGGCGG GCTCGTCGAC CTCGGCGCAG CGCTCGACGC GAGCGCCGCA 17940

GGCCGCTTGC TCCCGGCCCT CGCCCAGCGC CACGACGAAG ACCAGCTCGC GCTGCGCCCG 18000

GCCGGCCTCT ACGCACGCCG CTTCGTCCGC GCCCCGCTCG GCGATGCGCC TGCCGCTCGC 18060

GGCTTCATGC CCCGAGGCAC CATCCTCATC ACCGGTGGTA CCGGCGCCAT TGGCGCTCAC 18120

GTCGCCCGAT GGCTCGCTCG AAAAGGCGCT GAGCACCTCG TCCTCATCAG CCGACGAGGG 18180

GCCCAGGCCG AAGGCGCCGT GGAGCTCCAC GCCGAGCTCA CCGCCCTCGG CGCGCGCGTC 18240

ACCTTCGCCG CGTGCGATGT CGCCGACAGG AGCGCTGTCG CCACGCTTCT CGAGCAGCTC 18300

GACGCCGGAG GGCCACAGGT GAGCGCCGTG TTCCACGCGG GCGGCATCGA GCCCCACGCT 18360

CCGCTCGCCG CCACCTCCAT GGAGGATCTC GCCGAGGTTG TCTCCGGCAA GGTACAAGGT 18420

GCAAGACACC TCCACGACCT GCTCGGCTCT CGACCCCTCG ACGCCTTTGT TCTCTTCTCG 18480

TCCGGCGCGG TCGTCTGGGG CGGCGGACAA CAAGGCGGCT ATGCCGCTGC GAACGCCTTC 18540

CTCGATGCCC TGGCCGAGCA GCGGCGCAGC CTTGGGCTGA CGGCGACATC GGTGGCCTGG 18600

GGCGTGTGGG GCGGCGGCGG CATGGCTACC GGGCTCCTGG CAGCCCAGCT AGAGCAACGC 18660

GGTCTGTCGC CGATGGCCCC CTCGCTGGCC GTGGCGACGC TCGCGCTGGC GCTGGAGCAC 18720

GACGAGACCA CCCTCACCGT CGCCGACATC GACTGGGCGC GCTTTGCGCC TTCGTTCAGC 18780

GCCGCTCGCT CCCGCCCGCT CCTGCGCGAT TTGCCCGAGG CGCAGCGCGC TCTCGAAGCC 18840

AGCGCCGATG CGTCCTCCGA GCAAGACGGG GCCACAGGCC TCCTCGACAA GCTCCGAAAC 18900

CGCTCGGAGA GCGAGCAGAT CCACCTGCTC TCCTCGCTGG TGCGCCACGA AGCGGCCCTC 18960

GTCCTGGGCC ATACCGACGC CTCCCAGGTC GACCCCCACA AGGGCTTCAT GGACCTCGGC 19020

CTCGATTCGC TCATGACCGT CGAGCTTCGT CGGCGCTTGC AGCAGGCCAC CGGCATCAAG 19080

CTCCCGGCCA CCCTCGCCTT CGACCATCCC TCTCCTCATC GCGTCGCGCT CTTCTTGCGC 19140

GACTCGCTCG CCCACGCCCT CGGCGCGAGG CTCTCCGTCG AGCGCGACGC CGCCGCGCTC 19200

CCGGCGCTTC GCTCGGCGAG CGACGAGCCC ATCGCCATCG TCGGCATGGC CCTCCGCTTG 19260

CCGGGCGGCA TCGGCGATGT CGACGCTCTT TGGGAGTTCC TCGCCCAAGG ACGCGACGCC 19320

GTCGAGCCCA TTCCCCATGC CCGATGGGAT GCCGGTGCCC TCTACGACCC CGACCCCGAC 19380

GCCAAGGCCA AGAGCTACGT CCGGCATGCC GCCATGCTCG ACCAGGTCGA CCTCTTCGAT 19440

CCTGCCTTCT TTGGCATCAG CCCTCGCGAG GCCAAATACC TCGACCCCCA GCACCGCCTG 19500

CTCCTCGAAT CTGCCTGGCT GGCCCTCGAG GACGCCGGCA TCGTCCCCTC CACCCTCAAG 19560

GATTCTCCCA CCGGCGTCTT CGTCGGCATC GGCGCCAGCG AATACGCACT GCGAAACACG 19620

AGCTCCGAAG AGGTCGAAGC GTATGCCCTC CAAGGCACCG CCGGGTCCTT TGCCGCGGGG 19680

CGCTTGGCCT ACACGCTCGG CCTGCAAGGG CCCGCGCTCT CGGTCGACAC CGCCTGCTCC 19740

TCCTCGCTCG TCGCCCTCCA CCTCGCCTGC CAAGCCCTCC GACAGGGCGA GTGCAACCTC 19800

GCCCTCGCCG CGGGCGTCTC CGTCATGGCC TCCCCCGGGC TCTTCGTCGT CCTTTCCCGC 19860

ATGCGTGCTT TGGCGCCCGA TGGCCGCTCC AAGACCTTCT CGACCAACGC CGACGGCTAC 19920

GGACGCGGAG AGGGCGTCGT CGTCCTTGCC CTCGAGCGGC TCGGCGACGC CCTCGCCCGA 19980

GGACACCGCG TCCTCGCCCT CGTCCGCGGC ACCGCCATGA ACCATGACGG CGCGTCGAGC 20040

GGCATCACCG CCCCCAATGG CACCTCCCAC CAGAAGGTCC TCCGCGCCGC GCTCCACGAC 20100

GCCCATATCG GCCCTGCCGA CGTCGACGTC GTCGAATGCC ATGGCACCGG CACCTCCTTG 20160

GGAGACCCCA TCGAGGTGCA AGCCCTGGCC GCCGTCTACG CCGATGGCAG ACCCGCTGAA 20220

AAGCCTCTCC TTCTCGGCGC ACTCAAGACC AACATTGGCC ATCTCGAGGC CGCCTCCGGC 20280

CTCGCGGGCG TCGCCAAGAT CGTCGCCTCC CTCCGCCATG ACGCCCTGCC CCCCACCCTC 20340

CACACGACCC CGCGCAATCC CCTGATCGAG TGGGATGCGC TCGCCATCGA CGTCGTCGAT 20400

GCCACGAGGG CGTGGGCCCG CCACGAAGAT GGCAGTCCCC GCCGCGCCGG CGTCTCCGCC 20460
TTCGGACTCT CCGGCACCAA CGCCCACGTT ATCCTCGAAG AGGCTCCCGC GATCCCGCAG 20520
GCCGAGCCCA CCGCGGCACA GCTCGCGTCG CAGCCGCTTC CCGCAGCCTG GCCCGTGCTC 20580
CTGTCGGCCA GGAGCGAGCC GGCCGTGCGC GCCCAGGCCC AGAGGCTCCG CGACCACCTC 20640
CTCGCCCACG ACGACCTCGC CCTGGCCGAT GTAGCCTACT CGCTCGCCAC CACCCGGGCT 20700
ACCTTCGAGC ACCGTGCCGC TCTCGTGGTC CACGACCGCG AAGAGCTCCT CTCCGCGCTC 20760
GATTCGCTCG CCCAGGGAAG GCCCGCCCCG AGCACCGTCG TCGAACGAAG CGGAAGCCAC 20820
GGCAAGGTCG TCTTCGTCTT TCCTGGGCAA GGCTCGCAGT GGGAAGGGAT GGCCCTCTCC 20880
CTGCTCGATA CCTCGCCGGT CTTCCGGGCA CAGCTCGAAG CGTGCGAGCG CGCCCTCGCG 20940
CCCCACGTGG ACTGGTCGCT GCTCGCGGTG CTCCGCGGCG AGGAGGGCGC GCCCCCGCTC 21000
GACCGGGTCG ACGTGGTCCA GCCCGCGCTG TTCTCGATGA TGGTCTCGCT GGCCGCCCTG 21060
TGGCGCTCCA TGGGCGTCGA GCCCGACGCG GTGGTCGGCC ATAGCCAGGG CGAGATCGCC 21120
GCGGCCTGTG TGGCGGGCGC GCTGTCGCTC GAGGACGCTG CCAAGCTGGT GGCGCTGCGC 21180
AGCCGTGCGC TCGTGGAGCT CGCCGGCCAG GGGGCCATGG CCGCGGTGGA GCTGCCGGAG 21240
GCCGAGGTCG CACGGCGCCT CCAGCGCTAT GGCGATCGGC TCTCCATCGG GGCGATCAAC 21300
AGCCCTCGTT TCACGACGAT CTCCGGCGAG CCCCCTGCCG TCGCCGCCCT GCTCCGCGAT 21360
CTGGAGTCCG AGGGCGTCTT CGCCCTCAAG CTGAGTTACG ACTTCGCCTC CCACTCCGCG 21420
CAGGTCGAGT CGATTCGCGA CGAGCTCCTC GATCTCCTGT CGTGGCTCGA GCCGCGCTCG 21480
ACGGCGGTCC CGTTCTACTC CACGGTGAGC GGCGCCGCGA TCGACGGGAG CGAGCTCGAC 21540
GCCGCCTACT GGTACCGGAA CCTCCGGCAG CCGGTCCGCT TCGCAGACGC TGTGCAAGGC 21600
CTCCTTGCCG GAGAACATCG CTTCTTCGTG GAGGTGAGCC CCAGTCCTGT GCTGACCTTG 21660
GCCTTGCACG AGCTCCTCGA AGCGTCGGAG CGCTCGGCGG CGGTGGTCGG CTCTCTGTGG 21720
AGCGACGAAG GGGATCTACG GCGCTTCCTC GTCTCGCTCT CCGAGCTCTA CGTCAACGGC 21780
TTCGCCCTGG ATTGGACGAC GATCCTGCCC CCCGGGAAGC GGGTGCCGCT GCCCACCTAC 21840
CCCTTCCAGC GCGAGCGCTT CTGGCTCGAC GCCTCCACGG CACCCGCCGC CGGCGTCAAC 21900
CACCTTGCTC CGCTCGAGGG GCGGTTCTGG CAGGCCATCG AGAGCGGGAA TATCGACGCG 21960
CTCAGCGGCC AGCTCCACGT GGACGGCGAC GAGCAGCGCG CCGCCCTTGC CCTGCTCCTT 22020

CCCACCCTCG CGAGCTTTCG CCACGAGCGG CAAGAGCAGG GCACGGTCGA CGCCTGGCGC 22080
TACCGCATCA CGTGGAAGCC TCTGACCACC GCCACCACGC CCGCCGACCT GGCCGGCACC 22140
TGGCTCCTCG TCGTGCCGGC CGCTCTGGAC GACGACGCGC TCCCCTCCGC GCTCACCGAG 22200
GCGCTCGCCC GGCGCGGCGC GCGCGTCCTC GCCGTGCGCC TGAGCCAGGC CCACCTGGAC 22260
CGCGAGGCTC TCGCCGAGCA CCTGCGCCAG GCTTGCGCCG AGACCGCGCC GCCTCGCGGC 22320
GTGCTCTCGC TCCTCGCCCT CGACGAAAGT CCCCTCGCCG ACCATGCCGC CGTGCCCGCG 22380
GGACTCGCCT TCTCGCTCAC CCTCGTCCAA GCCCTCGGCG ACATCGCCCT CGACGCGCCC 22440
TTGTGGCTCT TCACCCGCGG CGCCGTCTCC GTCGGACACT CCGACCCCAT CGCCCATCCG 22500
ACGCAGGCGA TGACCTGGGG CCTGGGCCGC GTCGTCGGCC TCGAGCACCC CGAGCGCTGG 22560
GGAGGGCTCG TCGACGTCGG CGCAGCGATC GACGCGAGCG CCGTGGGCCG CTTGCTCCCG 22620
GTCCTCGCCC TGCGCAACGA TGAGGACCAG CTCGCTCTCC GCCCGGCCGG GTTCTACGCT 22680
CGCCGCCTCG TCCGCGCTCC GCTCGGCGAC GCGCCGCCCG CACGTACCTT CAAGCCCCGA 22740
GGCACCCTCC TCATCACCGG AGGCACCGGC GCCGCTGGCG CTCACGTCGC CCGATGGCTC 22800
GCTCGAGAAG GCGCAGAGCA CCTCGTCCTC ATCAGCCGCC GAGGGGCCCA GGCCGAGGGC 22860
GCCTCGGAGC TCCACGCCGA GCTCACGGCC CTGGGCGCGC GCGTCACCTT CGCCGCGTGT 22920
GATGTCGCCG ACAGGAGCGC TGTCGCCACG CTTCTCGAGC AGCTCGACGC CGAAGGGTCG 22980
CAGGTCCGCG CCGTGTTCCA CGCGGGCGGC ATCGGGCGCC ACGCTCCGCT CGCCGCCACC 23040
TCTCTCATGG AGCTCGCCGA CGTTGTCTCT GCCAAGGTCC TAGGCGCAGG GAACCTCCAC 23100
GACCTGCTCG GTCCTCGACC CCTCGACGCC TTCGTCCTTT TCTCGTCCAT CGCAGGCGTC 23160
TGGGGCGGCG GACAACAAGC CGGATACGCC GCCGGAAACG CCTTCCTCGA CGCCCTGGCC 23220
GACCAGCGGC GCAGTCTTGG ACAGCCGGAC ACGTCCGTGG TGTGGGGCGC GTGGGGCGGC 23280
GGCGGTGGTA TATTCACGGG GCCCCTGGCA GCCCAGCTGG AGCAACGTCG TCTGTCGCCG 23340
ATGGCCCCTT CGCTGGCCGT GGCGGCGCTC GCGCAAGCCC TGGAGCACGA CGAGACCACC 23400
GTCACCGTCG CCGACATCGA CTGGGCGCGC TTTGCGCCTT CGATCAGCGT CGCTCGCTCC 23460
CGCCGCTCCT GCGCGACTTG CCCGAGCAGC GCGCCCTCGA AGACAGAGAA GGCGCGTCCT 23520
CCTCCGAGCA CGGCCCGGCC CCCCGACCTC CTCGACAAGC TCCGGAGCCG CTCGGAGAGC 23580
GAGCAGCTCC GTCTGCTCGC CGCGCTGGTG TGCGACGAGA CGGCCCTCGT CCTCGGCCAC 23640
GAAGGCCGCT TCCCAGCTCG ACCCCGACAA GGCTTCTTCG ACCTCGGTCT CGATTCGATC 23700

ATGACCGTCG AGCTTCGTCG GCGCTTGCAA CAGGCCACCG GCATCAAGCT CCCGGCCACC 23760

CTCGCCTTCG ACCATCCCTC TCCTCATCGC GTCGCGCTCT TCATGCGCGA CTCGCTCGCC 23820

CACGCCCTCG GCACGAGGCT CTCCGCCGAG GCGACGCCGC CGCGCTCCGG CCGCGCCTCG 23880

AGCGACGAGC CCATCGCCAT CGTCGGCATG GCCCTGCGCC TGCCGGGCGG CGTCGGCGAT 23940

GTCGACGCTC TTTGGGAGTT CCTCCACCAA GGGCGCGACG CGGTCGAGCC CATTCCACAG 24000

AGCCGCTGGG ACGCCGGTGC CCTCTACGAC CCCGACCCCG ACGCCGACGC CAAGAGCTAC 24060

GTCCGGCATG CCGCGATGCT CGACCAGATC GACCTCTTCG ACCCTGCCTT CTTCGGCATC 24120

AGCCCCCGGG AGGCCAAACA CCTCGACCCC CAGCACCGCC TGCTCCTCGA ATCTGCCTGG 24180

CTGGCCCTCG AGGACGCCGG CATCGTCCCC ACCTCCCTCA AGGACTCCCT CACCGGCGTC 24240

TTCGTCGGCA TCTGCGCCGG CGAATACGCG ATGCAAGAGG CGAGCTCGGA AGGTTCCGAG 24300

GTTTACTTCA TCCAAGGCAC TTCCGCGTCC TTTGGCGCGG GGGGCTTGGC CTATACGCTC 24360

GGGCTCCAGG GGCCGCGATC TTCGGTCGAC ACCGCCTGCT CCTCCTCGCT CGTCTCCCTC 24420

CACCTCGCCT GCCAAGCCCT CCGACAGGGC GAGTGCAACC TCGCCCTCGC CGCGGGCGTG 24480

TCGCTCATGG TCTCCCCCCA GACCTTCGTC ATCCTTTCCC GTCTGCGCGC CTTGGCGCCC 24540

GACGGCCGCT CCAAGACCTT CTCGGACAAC GCCGACGGCT ACGGACGCGG AGAAGGCGTC 24600

GTCGTCCTTG CCCTCGAGCG GATCGGCGAC GCCCTCGCCC GGAGACACCG CGTCCTCGTC 24660

CTCGTCCGCG GCACCGCCAT CAACCACGAC GGCGCGTCGA GCGGTATCAC CGCCCCCAAC 24720

GGCACCTCCC AGCAGAAGGT CCTCCGGGCC GCGCTCCACG ACGCCCGCAT CACCCCCGCC 24780

GACGTCGACG TCGTCGAGTG CCATGGCACC GGCACCTCGC TGGGAGACCC CATCGAGGTG 24840

CAAGCCCTGG CCGCCGTCTA CGCCGACGGC AGACCCGCTG AAAAGCCTCT CCTTCTCGGC 24900

GCGCTCAAGA CCAACATCGG CCATCTCGAG GCCGCCTCCG GCCTCGCGGG CGTCGCCAAG 24960

ATGGTCGCCT CGCTCCGCCA CGACGCCCTG CCCCCCACCC TCCACGCGAC CCCACGCAAT 25020

CCCCTCATCG AGTGGGAGGC GCTCGCCATC GACGTCGTCG ATACCCCGAG GCCTTGGCCC 25080

CGCCACGAAG ATGGCAGTCC CCGCCGCGCC GGCATCTCCG CCTTCGGATT CTCGGGCACC 25140

AACGCCCACG TCATCCTCGA AGAGGCTCCC GCCGCCCTGC CGGCCGAGCC CGCCACCTCA 25200

CAGCCGGCGT CGCAAGCCGC TCCCGCGGCG TGGCCCGTGC TCCTGTCGGC CAGGAGCGAG 25260

GCCGCCGTCC GCGCCCAGGC GAAGCGGCTC CGCGACCACC TCGTCGCCCA CGACGACCTC 25320

ACCCTCGCGG ATGTGGCCTA TTCGCTGGCC ACCACCCGCG CCCACTTCGA GCACCGCGCC 25380

GCTCTCGTAG CCCACAACCG CGACGAGCTC CTCTCCGCGC TCGACTCGCT CGCCCAGGAC 25440

AAGCCCGCCC CGAGCACCGT CCTCGGACGG AGCGGAAGCC ACGGCAAGCT CGTCTTCGTC 25500

TTTCCTGGGC AAGGCTCGCA GTGGGAAGGG ATGGCCCTCT CGCTGCTCGA CTCCTCGCCC 25560

GTCTTCCGCG CTCAGCTCGA AGCATGCGAG CGCGCGCTCG CTCCTCACGT CGAGTGGAGC 25620

CTGCTCGCCG TCCTGCGCCG CGACGAGGGC GCCCCCTCCC TCGACCGCGT CGACGTCGTA 25680

CAGCCCGCCC TCTTTGCCGT CATGGTCTCC CTGGCGGCCC TCTGGCGCTC GCTCGGCGTA 25740

GAGCCCGCCG CCGTCGTCGG CCACAGTCAG GGCGAGATCG CCGCCGCCTT CGTCGCAGGC 25800

GCTCTCTCCC TCGAGGACGC GGCCCGCATC GCCGCCCTGC GCAGCAAAGC GCTCACCACC 25860

GTCGCCGGCA ACGGGGCCAT GGCCGCCGTC GAGCTCGGCG CCTCCGACCT CCAGACCTAC 25920

CTCGCTCCCT GGGGCGACAG GCTCTCCATC GCCGCCGTCA ACAGCCCCAG GGCCACGCTC 25980

GTGTCCGGCG AGCCCGCCGC CATCGACGCG CTGATCGACT CGCTCACCGC AGCGCAGGTC 26040

TTCGCCCGAA AAGTCCGCGT CGACTACGCC TCCCACTCCG CCCAGATGGA CGCCGTCCAA 26100

GACGAGCTCG CCGCAGGTCT AGCCAACATC GCTCCTCGGA CGTGCGAGCT CCCTCTTTAT 26160

TCGACCGTCA CCGGCACCAG GCTCGACGGC TCCGAGCTCG ACGGCGCGTA CTGGTATCGA 26220

AACCTCCGGC AAACCGTCCT GTTCTCGAGC GCGACCGAGC GGCTCCTCGA CGATGGGCAT 26280

CGCTTCTTCG TCGAGGTCAG CCCCCATCCC GTGCTCACGC TCGCCCTCCG CGAGACCTGC 26340

GAGCGCTCAC CGCTCGATCC CGTCGTCGTC GGCTCCATTC GACGCGACGA AGGCCACCTC 26400

GCCCGCCTGC TCCTCTCCTG GGCGGAGCTC TCTACCCGAG GCCTCGCGCT CGACTGGAAC 26460

GCCTTCTTCG CGCCCTTCGC TCCCCGCAAG GTCTCCCTCC CCACCTACCC CTTCCAACGC 26520

GAGCGCTTCT GGCTCGACGC CTCCACGGCG CACGCTGCCG ACGTCGCCTC CGCAGGCCTG 26580

ACCTCGGCCG ACCACCCGCT GCTCGGCGCC GCCGTCGCCC TCGCCGACCG CGATGGCTTT 26640

GTCTTCACAG GACGGCTCTC CCTCGCAGAG CACCCGTGGC TCGAAGACCA CGTCGTCTTC 26700

GGCATACCCT GTCCTGCCAG GCGCCGCCTC CTCGAGCTCG CCCTGCATGT CGCCCATCTC 26760

GTCGGCCTCG ACACCGTCGA AGACGTCACG CTCGACCCCC CCCTCGCTCT CCCATCGCAG 26820

GGCGCCGTCC TCCTCCAGAT CTCCGTCGGG CCCGCGGACG GTGCTGGACG AAGGGCGCTC 26880

TCCGTTCATA GCCGGCGCCA CGACGCGCTT CAGGATGGCC CCTGGACTCG CCACGCCAGC 26940

GGCTCTCTCG CGCAAGCTAG CCCGTCCCAT TGCCTTCGAT GCTCCGCGAA TGGCCCCCCC 27000

TCGGGCGCCA CCCAGGTGGA CACCCAAGGT TTCTACGCAG CCCTCGAGAG CGCTGGGCTT 27060
GCTTATGGCC CCGAGTTCCA GGGCCTCCGC CGCCGTCTAC AAGCGCGGCG ACGAGCTCTT 27120
CGCCGAAGCC AAGCTCCCGG ACGCCGCCGA AGAGGACGCC GCTCGTTTTG CCCTCCACCC 27180
CGCCCTGCTC GACAGCGCCT TGCAGGCGCT CGCCTTTGTA GACGACCAGG CAAAGGCCTT 27240
CAGGATGCCC TTCTCGTGGA GCGGAGTATC GCTGCGCTCC GGTCGGAGCC ACCACCCTGC 27300
GCGTGCGTTT CCACCGTCCT GAGGGCGAAT CCTCGCGCTC GCTCCTCCTC GCCGACGCCA 27360
GAGGCGAACC CATCGCCTCG GTGCAAGCGC TCGCCATGCG CGCCGCGTCC GCCGAGCAGC 27420
TCCGCAGACC CGGGAGCGTC CCACCTCGAT GCCCTCTTCC GCATCGACTG GAGCGAGCTG 27480
CAAAGCCCCA CCTCACCGCC CATCGCCCCG AGCGGTGCCC TCCTCGGCAC AGAAGGTCTC 27540
GACCTCGGGA CCAGGGTGCC TCTCGACCGC TATACCGACC TTGCTGCTCT ACGCAGCGCC 27600
CTCGACCAGG GCGCTTCGCC TCCAAGCCTC GTCATCGCCC CCTTCATCGC TCTGCCCGAA 27660
GGCGACCTCA TCGCGAGCGC CCGCGAGACC ACCGCGCACG CGCTCGCCCT CTTGCAAGCC 27720
TGGCTCGCCG ACGAGCGCCT CGCCTCCTCG CGCCTCGCCC TCGTCACCCG ACGCGCCGTC 27780
GCCACCCACG CTGAAGAAGA CGTCAAGGGC CTCGCTCACG CGCCTCTCTG GGGTCTCGCT 27840
CGCTCCGCGC AGAGCGAGCA CCCAGAGCGC CCTCTCGTCC TCGTCGACCT CGACGACAGC 27900
GAGGCCTCCC AGCACGCCCT GCTCGGCGCG CTCGACGCAA GAGAGCCAGA GATCGCCCTC 27960
CGCAACGGCA AACCCCTCGT TCCAAGGCTC TCACGCCTGC CCCAGGCGCC CACGGACACA 28020
GCGTCCCCCG CAGGCCTCGG AGGCACCGTC CTCATCACGG GAGGCACCGG CACGCTCGGC 28080
GCCCTGGTCG CGCGCCGCCT CGTCGTAAAC CACGACGCCA AGCACCTGCT CCTCACCTCG 28140
CGCCAGGGCG CGAGCGCTCC GGGTGCTGAT GTCTTGCGAA GCGAGCTCGA AGCTCTGGGG 28200
GCTTCGGTCA CCCTCGCCGC GTGCGACGTG GCCGATCCAC GCGCTCTAAA GGACCTTCTG 28260
GATAACATTC CGAGCGCTCA CCCGGTCGCC GCCGTCGTGC ATGCCGCCAG CGTCCTCGAC 28320
GGCGATCTGC TCGGCGCCAT GAGCCTCGAG CGGATCGACC GCGTCTTCGC CCCCAAGATC 28380
GATGCCGCCT GGCACTTGCA TCAGCTCACC CAAGATAAGC CCCTTGCCGC CTTCATCCTC 28440
TTCTCGTCCG TCGCCGGCGT CCTCGGCAGC TCAGGTCACT CCAACTACGC CGCTGCGAGC 28500
GCCTTCCTCG ATGCGCTTGC GCACCACCGG CGCGCGCAAG GGCTCCCTGC CTCATCGCTC 28560
GCGTGGAGCC ACTGGGCCGA GCGCAGCGCA ATGACAGAGC ACGTCAGCGC CGCCGGCGCC 28620

CCTCGCATGG AGCGCGCCGG CCTTCCCTCG ACCTCTGAGG AGAGGCTCGC CCTCTTCGAT 28680

GCGGCGCTCT TCCGAACCGA GACCGCCCTG GTCCCCGCGC GCTTCGACTT GAGCGCGCTC 28740

AGGGCGAACG CCGGCAGCGT CCCCCCGTTG TTCCAACGTC TCGTCCGCGC TCGCACCGTA 28800

CGCAAGGCCG CCAGCAACAC CGCCCAGGCC TCGTCGCTTA CAGAGCGCCT CTCAGCCCTC 28860

CCGCCCGCCG AACGCGAGCG TGCCCTGCTC GATCTCATCC GCACCGAAGC CGCCGCCGTC 28920

CTCGGCCTCG CCTCCTTCGA ATCGCTCGAT CCCGATCG 28958

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..13
(D) OTHER INFORMATION: /note= "sequence of a plant consensus translation initiator (Clontech)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GTCGACCATG GTC 13

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature

(B) LOCATION: 1..12
(D) OTHER INFORMATION: /note= "sequence of a plant consensus translation initiator (Joshi)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TAAACAATGG CT 12

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..22
(D) OTHER INFORMATION: /note= "sequence of an oligonucleotide for use in a molecular adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AATTCTAAAG CATGCCGATC GG 22

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..21
(D) OTHER INFORMATION: /note= "sequence of an oligonucleotide for use in a molecular adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AATTCCGATC GGCATGCTTT A 21

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..22
(D) OTHER INFORMATION: /note= "sequence of an oligonucleotide for use in a molecular adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AATTCTAAAC CATGGCGATC GG 22

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..21
(D) OTHER INFORMATION: /note= "sequence of an oligonucleotide for use in a molecular adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AATTCCGATC GCCATGGTTT A 21

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..15
(D) OTHER INFORMATION: /note= "sequence of an oligonucleotide for use in a molecular adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCAGCTGGAA TTCCG 15

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..19
(D) OTHER INFORMATION: /note= "sequence of an oligonucleotide for use in a molecular adaptor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CGGAATTCCA GCTGGCATG 19

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..11
(D) OTHER INFORMATION: /note= "oligonucleotide used to introduce base change into SphI site of ORF1 of pyrrolnitrin gene cluster"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCCCCTCATG C 11

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..11
(D) OTHER INFORMATION: /note= "oligonucleotide used to introduce base change into SphI site of ORF1 of pyrrolnitrin gene cluster"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GCATGAGGGG G 11

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4603 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 230..1597
(D) OTHER INFORMATION: /gene= "phz1"
/label= ORF1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1598..2761
(D) OTHER INFORMATION: /gene= "phz2"
/label= ORF2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2764..3600
(D) OTHER INFORMATION: /gene= "phz3"
/label= ORF3

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 3597..4265
(D) OTHER INFORMATION: /label= ORF4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
GCATGCCGTG ACCTCCGCCG GTGGCGTGGC CGCCGGCCTG CACCTGGAAA CCACCCCTGA      60

CGACGTCAGC GAGTGCGCTT CCGATGCCGC CGGCCTGCAT CAGGTCGCCA GCCGCTACAA     120

AAGCCTGTGC GACCCGCGCC TGAACCCCTG GCAAGCCATT ACTGCGGTGA TGGCCTGGAA     180

AAACCAGCCC TCTTCAACCC TTGCCTCCTT TTGACTGGAG TTTGTCGTC ATG ACC         235
                                                      Met Thr
                                                        1

GGC ATT CCA TCG ATC GTC CCT TAC GCC TTG CCT ACC AAC CGC GAC CTG       283
Gly Ile Pro Ser Ile Val Pro Tyr Ala Leu Pro Thr Asn Arg Asp Leu
        5                  10                  15

CCC GTC AAC CTC GCG CAA TGG AGC ATC GAC CCC GAG CGT GCC GTG CTG       331
```

```
Pro Val Asn Leu Ala Gln Trp Ser Ile Asp Pro Glu Arg Ala Val Leu
     20                  25                  30

CTG GTG CAT GAC ATG CAG CGC TAC TTC CTG CGG CCC TTG CCC GAC GCC      379
Leu Val His Asp Met Gln Arg Tyr Phe Leu Arg Pro Leu Pro Asp Ala
 35                  40                  45                  50

CTG CGT GAC GAA GTC GTG AGC AAT GCC GCG CGC ATT CGC CAG TGG GCT      427
Leu Arg Asp Glu Val Val Ser Asn Ala Ala Arg Ile Arg Gln Trp Ala
                 55                  60                  65

GCC GAC AAC GGC GTT CCG GTG GCC TAC ACC GCC CAG CCC GGC AGC ATG      475
Ala Asp Asn Gly Val Pro Val Ala Tyr Thr Ala Gln Pro Gly Ser Met
             70                  75                  80

AGC GAG GAG CAA CGC GGG CTG CTC AAG GAC TTC TGG GCC CCG GGC ATG      523
Ser Glu Glu Gln Arg Gly Leu Leu Lys Asp Phe Trp Gly Pro Gly Met
         85                  90                  95

AAG GCC AGC CCC GCC GAC CGC GAG GTG GTC GGC GCC CTG ACG CCC AAG      571
Lys Ala Ser Pro Ala Asp Arg Glu Val Val Gly Ala Leu Thr Pro Lys
    100                 105                 110

CCC GGC GAC TGG CTG CTG ACC AAG TGG CGC TAC AGC GCG TTC TTC AAC      619
Pro Gly Asp Trp Leu Leu Thr Lys Trp Arg Tyr Ser Ala Phe Phe Asn
115                 120                 125                 130

TCC GAC CTG CTG GAA CGC ATG CGC GCC AAC GGG CGC GAT CAG TTG ATC      667
Ser Asp Leu Leu Glu Arg Met Arg Ala Asn Gly Arg Asp Gln Leu Ile
                135                 140                 145

CTG TGC GGG GTG TAC GCC CAT GTC GGG GTA CTG ATT TCC ACC GTG GAT      715
Leu Cys Gly Val Tyr Ala His Val Gly Val Leu Ile Ser Thr Val Asp
            150                 155                 160

GCC TAC TCC AAC GAT ATC CAG CCG TTC CTC GTT GCC GAC GCG ATC GCC      763
Ala Tyr Ser Asn Asp Ile Gln Pro Phe Leu Val Ala Asp Ala Ile Ala
        165                 170                 175

GAC TTC AGC AAA GAG CAC CAC TGG ATG CCA TCG AAT ACG CCG CCA GCC      811
Asp Phe Ser Lys Glu His His Trp Met Pro Ser Asn Thr Pro Pro Ala
    180                 185                 190

GTT GCG CCA TGT CAT CAC CAC CGA CGA GGT GGT GCT ATG AGC CAG ACC      859
Val Ala Pro Cys His His His Arg Arg Gly Gly Ala Met Ser Gln Thr
195                 200                 205                 210

GCA GCC CAC CTC ATG GAA CGC ATC CTG CAA CCG GCT CCC GAG CCG TTT      907
Ala Ala His Leu Met Glu Arg Ile Leu Gln Pro Ala Pro Glu Pro Phe
                215                 220                 225

GCC CTG TTG TAC CGC CCG GAA TCC AGT GGC CCC GGC CTG CTG GAC GTG      955
Ala Leu Leu Tyr Arg Pro Glu Ser Ser Gly Pro Gly Leu Leu Asp Val
            230                 235                 240
```

```
CTG ATC GGC GAA ATG TCG GAA CCG CAG GTC CTG GCC GAT ATC GAC TTG    1003
Leu Ile Gly Glu Met Ser Glu Pro Gln Val Leu Ala Asp Ile Asp Leu
        245                 250                 255

CCT GCC ACC TCG ATC GGC GCG CCT CGC CTG GAT GTA CTG GCG CTG ATC    1051
Pro Ala Thr Ser Ile Gly Ala Pro Arg Leu Asp Val Leu Ala Leu Ile
    260                 265                 270

CCC TAC CGC CAG ATC GCC GAA CGC GGT TTC GAG GCG GTG GAC GAT GAG    1099
Pro Tyr Arg Gln Ile Ala Glu Arg Gly Phe Glu Ala Val Asp Asp Glu
275                 280                 285                 290

TCG CCG CTG CTG GCG ATG AAC ATC ACC GAG CAG CAA TCC ATC AGC ATC    1147
Ser Pro Leu Leu Ala Met Asn Ile Thr Glu Gln Gln Ser Ile Ser Ile
                295                 300                 305

GAG CGC TTG CTG GGA ATG CTG CCC AAC GTG CCG ATC CAG TTG AAC AGC    1195
Glu Arg Leu Leu Gly Met Leu Pro Asn Val Pro Ile Gln Leu Asn Ser
            310                 315                 320

GAA CGC TTC GAC CTC AGC GAC GCG AGC TAC GCC GAG ATC GTC AGC CAG    1243
Glu Arg Phe Asp Leu Ser Asp Ala Ser Tyr Ala Glu Ile Val Ser Gln
        325                 330                 335

GTG ATC GCC AAT GAA ATC GGC TCC GGG GAA GGC GCC AAC TTC GTC ATC    1291
Val Ile Ala Asn Glu Ile Gly Ser Gly Glu Gly Ala Asn Phe Val Ile
    340                 345                 350

AAA CGC ACC TTC CTG GCC GAG ATC AGC GAA TAC GGC CCG GCC AGT GCG    1339
Lys Arg Thr Phe Leu Ala Glu Ile Ser Glu Tyr Gly Pro Ala Ser Ala
355                 360                 365                 370

CTG TCG TTC TTT CGC CAT CTG CTG GAA CGG GAG AAA GGC GCC TAC TGG    1387
Leu Ser Phe Phe Arg His Leu Leu Glu Arg Glu Lys Gly Ala Tyr Trp
                375                 380                 385

ACG TTC ATC ATC CAC ACC GGC AGC CGT ACC TTC GTG GGT GCG TCC CCC    1435
Thr Phe Ile Ile His Thr Gly Ser Arg Thr Phe Val Gly Ala Ser Pro
            390                 395                 400

GAG CGC CAC ATC AGC ATC AAG GAT GGG CTC TCG GTG ATG AAC CCC ATC    1483
Glu Arg His Ile Ser Ile Lys Asp Gly Leu Ser Val Met Asn Pro Ile
        405                 410                 415

AGC GGC ACT TAC CGC TAT CCG CCC GCC GGC CCC AAC CTG TCG GAA GTC    1531
Ser Gly Thr Tyr Arg Tyr Pro Pro Ala Gly Pro Asn Leu Ser Glu Val
    420                 425                 430

ATG GAC TTC CTG GCG GAT CGC AAG GAA GCC GAC GAG CTC TAC ATG GTG    1579
Met Asp Phe Leu Ala Asp Arg Lys Glu Ala Asp Glu Leu Tyr Met Val
435                 440                 445                 450

GTG GAT GAA GAG CTG TAA ATG ATG GCG CGC ATT TGT GAG GAC GGC GGC    1627
Val Asp Glu Glu Leu  *  Met Met Ala Arg Ile Cys Glu Asp Gly Gly
                455      1                   5                  10
```

```
CAC GTC CTC GGC CCT TAC CTC AAG GAA ATG GCG CAC CTG GCC CAC ACC      1675
His Val Leu Gly Pro Tyr Leu Lys Glu Met Ala His Leu Ala His Thr
                 15                  20                  25

GAG TAC TTC ATC GAA GGC AAG ACC CAT CGC GAT GTA CGG GAA ATC CTG      1723
Glu Tyr Phe Ile Glu Gly Lys Thr His Arg Asp Val Arg Glu Ile Leu
             30                  35                  40

CGC GAA ACC CTG TTT GCG CCC ACC GTC ACC GGC AGC CCA CTG GAA AGC      1771
Arg Glu Thr Leu Phe Ala Pro Thr Val Thr Gly Ser Pro Leu Glu Ser
         45                  50                  55

GCC TGC CGG GTC ATC CAG CGC TAT GAN CCG CAA GGC CGC GCG TAC TAC      1819
Ala Cys Arg Val Ile Gln Arg Tyr Xaa Pro Gln Gly Arg Ala Tyr Tyr
     60                  65                  70

AGC GGC ATG GCT GCG CTG ATC GGC AGC GAT GGC AAG GGC GGG CGT TCC      1867
Ser Gly Met Ala Ala Leu Ile Gly Ser Asp Gly Lys Gly Gly Arg Ser
 75                  80                  85                  90

CTG GAC TCC GCG ATC CTG ATT CGT ACC GCC GAC ATC GAT AAC AGC GGC      1915
Leu Asp Ser Ala Ile Leu Ile Arg Thr Ala Asp Ile Asp Asn Ser Gly
                 95                 100                 105

GAG GTG CGG ATC AGC GTG GGC TCG ACC ATC GTG CGC CAT TCC GAC CCG      1963
Glu Val Arg Ile Ser Val Gly Ser Thr Ile Val Arg His Ser Asp Pro
            110                 115                 120

ATG ACC GAG GCT GCC GAA AGC CGG GCC AAG GCC ACT GGC CTG ATC AGC      2011
Met Thr Glu Ala Ala Glu Ser Arg Ala Lys Ala Thr Gly Leu Ile Ser
        125                 130                 135

GCA CTG AAA AAC CAG GCG CCC TCG CGC TTC GGC AAT CAC CTG CAA GTG      2059
Ala Leu Lys Asn Gln Ala Pro Ser Arg Phe Gly Asn His Leu Gln Val
    140                 145                 150

CGC GCC GCA TTG GCC AGC CGC AAT GCC TAC GTC TCG GAC TTC TGG CTG      2107
Arg Ala Ala Leu Ala Ser Arg Asn Ala Tyr Val Ser Asp Phe Trp Leu
155                 160                 165                 170

ATG GAC AGC CAG CAG CGG GAG CAG ATC CAG GCC GAC TTC AGT GGG CGC      2155
Met Asp Ser Gln Gln Arg Glu Gln Ile Gln Ala Asp Phe Ser Gly Arg
                175                 180                 185

CAG GTG CTG ATC GTC GAC GCC GAA GAC ACC TTC ACC TCG ATG ATC GCC      2203
Gln Val Leu Ile Val Asp Ala Glu Asp Thr Phe Thr Ser Met Ile Ala
            190                 195                 200

AAG CAA CTG CGG GCC CTG GGC CTG GTA GTG ACG GTG TGC AGC TTC AGC      2251
Lys Gln Leu Arg Ala Leu Gly Leu Val Val Thr Val Cys Ser Phe Ser
        205                 210                 215

GAC GAA TAC AGC TTT GAA GGC TAC GAC CTG GTC ATC ATG GGC CCC GGC      2299
Asp Glu Tyr Ser Phe Glu Gly Tyr Asp Leu Val Ile Met Gly Pro Gly
```

```
    220                 225                 230

CCC GGC AAC CCG AGC GAA GTC CAA CAG CCG AAA ATC AAC CAC CTG CAC        2347
Pro Gly Asn Pro Ser Glu Val Gln Gln Pro Lys Ile Asn His Leu His
235                 240                 245                 250

GTG GCC ATC CGC TCC TTG CTC AGC CAG CAG CGG CCA TTC CTC GCG GTG        2395
Val Ala Ile Arg Ser Leu Leu Ser Gln Gln Arg Pro Phe Leu Ala Val
                255                 260                 265

TGC CTG AGC CAT CAG GTG CTG AGC CTG TGC CTG GGC CTG GAA CTG CAG        2443
Cys Leu Ser His Gln Val Leu Ser Leu Cys Leu Gly Leu Glu Leu Gln
            270                 275                 280

CGC AAA GCC ATT CCC AAC CAG GGC GTG CAA AAA CAG ATC GAC CTG TTT        2491
Arg Lys Ala Ile Pro Asn Gln Gly Val Gln Lys Gln Ile Asp Leu Phe
        285                 290                 295

GGC AAT GTC GAA CGG GTG GGT TTC TAC AAC ACC TTC GCC GCC CAG AGC        2539
Gly Asn Val Glu Arg Val Gly Phe Tyr Asn Thr Phe Ala Ala Gln Ser
    300                 305                 310

TCG AGT GAC CGC CTG GAC ATC GAC GGC ATC GGC ACC GTC GAA ATC AGC        2587
Ser Ser Asp Arg Leu Asp Ile Asp Gly Ile Gly Thr Val Glu Ile Ser
315                 320                 325                 330

CGC GAC AGC GAG ACC GGC GAG GTG CAT GCC CTG CGT GGC CCC TCG TTC        2635
Arg Asp Ser Glu Thr Gly Glu Val His Ala Leu Arg Gly Pro Ser Phe
                335                 340                 345

GCC TCC ATG CAG TTT CAT GCC GAG TCG CTG CTG ACC CAG GAA GGT CCG        2683
Ala Ser Met Gln Phe His Ala Glu Ser Leu Leu Thr Gln Glu Gly Pro
            350                 355                 360

CGC ATC ATC GCC GAC CTG CTG CGG CAC GCC CTG ATC CAC ACA CCT GTC        2731
Arg Ile Ile Ala Asp Leu Leu Arg His Ala Leu Ile His Thr Pro Val
        365                 370                 375

GAG AAC AAC GCT TCG GCC GCC GGG AGA TAA CC ATG CAC CAT TAC GTC         2778
Glu Asn Asn Ala Ser Ala Ala Gly Arg  *     Met His His Tyr Val
    380                 385                   1               5

ATC ATC GAC GCC TTT GCC AGC GTC CCG CTG GAA GGC AAT CCG GTC GCG        2826
Ile Ile Asp Ala Phe Ala Ser Val Pro Leu Glu Gly Asn Pro Val Ala
                10                  15                  20

GTG TTC TTT GAC GCC GAT GAC TTG TCG GCC GAG CAA ATG CAA CGC ATT        2874
Val Phe Phe Asp Ala Asp Asp Leu Ser Ala Glu Gln Met Gln Arg Ile
            25                  30                  35

GCC CGG GAG ATG AAC CTG TCG GAA ACC ACT TTC GTG CTC AAG CCA CGT        2922
Ala Arg Glu Met Asn Leu Ser Glu Thr Thr Phe Val Leu Lys Pro Arg
        40                  45                  50

AAC TGC GGC GAT GCG CTG ATC CGG ATC TTC ACC CCG GTC AAC GAA CTG        2970
```

Asn Cys Gly Asp Ala Leu Ile Arg Ile Phe Thr Pro Val Asn Glu Leu
55 60 65

CCC TTC GCC GGG CAC CCG TTG CTG GGC ACG GAC ATT GCC CTG GGT GCG 3018
Pro Phe Ala Gly His Pro Leu Leu Gly Thr Asp Ile Ala Leu Gly Ala
70 75 80 85

CGC ACC GAC AAT CAC CGG CTG TTC CTG GAA ACC CAG ATG GGC ACC ATC 3066
Arg Thr Asp Asn His Arg Leu Phe Leu Glu Thr Gln Met Gly Thr Ile
90 95 100

GCC TTT GAG CTG GAG CGC CAG AAC GGC AGC GTC ATC GCC GCC AGC ATG 3114
Ala Phe Glu Leu Glu Arg Gln Asn Gly Ser Val Ile Ala Ala Ser Met
105 110 115

GAC CAG CCG ATA CCG ACC TGG ACG GCC CTG GGG CGC GAC GCC GAG TTG 3162
Asp Gln Pro Ile Pro Thr Trp Thr Ala Leu Gly Arg Asp Ala Glu Leu
120 125 130

CTC AAG GCC CTG GGC ATC AGC GAC TCG ACC TTT CCC ATC GAG ATC TAT 3210
Leu Lys Ala Leu Gly Ile Ser Asp Ser Thr Phe Pro Ile Glu Ile Tyr
135 140 145

CAC AAC GGC CCG CGT CAT GTG TTT GTC GGC CTG CCA AGC ATC GCC GCG 3258
His Asn Gly Pro Arg His Val Phe Val Gly Leu Pro Ser Ile Ala Ala
150 155 160 165

CTG TCG GCC CTG CAC CCC GAC CAC CGT GCC CTG TAC AGC TTC CAC GAC 3306
Leu Ser Ala Leu His Pro Asp His Arg Ala Leu Tyr Ser Phe His Asp
170 175 180

ATG GCC ATC AAC TGT TTT GCC GGT GCG GGA CGG CGC TGG CGC AGC CGG 3354
Met Ala Ile Asn Cys Phe Ala Gly Ala Gly Arg Arg Trp Arg Ser Arg
185 190 195

ATG TTC TCG CCG GCC TAT GGG GTG GTC GAG GAT GCG NCC ACG GGC TCC 3402
Met Phe Ser Pro Ala Tyr Gly Val Val Glu Asp Ala Xaa Thr Gly Ser
200 205 210

GCT GCC GGG CCC TTG GCG ATC CAT CTG GCG CGG CAT GGC CAG ATC GAG 3450
Ala Ala Gly Pro Leu Ala Ile His Leu Ala Arg His Gly Gln Ile Glu
215 220 225

TTC GGC CAG CAG ATC GAA ATT CTT CAG GGC GTG GAA ATC GGC CGC CCC 3498
Phe Gly Gln Gln Ile Glu Ile Leu Gln Gly Val Glu Ile Gly Arg Pro
230 235 240 245

TCA CTC ATG TTC GCC CGG GCC GAG GGC CGC GCC GAT CAA CTG ACG CGG 3546
Ser Leu Met Phe Ala Arg Ala Glu Gly Arg Ala Asp Gln Leu Thr Arg
250 255 260

GTC GAA GTA TCA GGC AAT GGC ATC ACC TTC GGA CGG GGG ACC ATC GTT 3594
Val Glu Val Ser Gly Asn Gly Ile Thr Phe Gly Arg Gly Thr Ile Val
265 270 275

CTA TGA ACAGTTCAGT ACTAGGCAAG CCGCTGTTGG GTAAAGGCAT GTCGGAATCG 3650
Leu *

CTGACCGGCA CACTGGATGC GCCGTTCCCC GAGTACCAGA AGCCGCCTGC CGATCCCATG 3710

AGCGTGCTGC ACAACTGGCT CGAACGCGCA CGCCGCGTGG GCATCCGCGA ACCCCGTGCG 3770

CTGGCGCTGG CCACGGCTGA CAGCCAGGGC CGGCCTTCGA CACGCATCGT GGTGATCAGT 3830

GAGATCAGTG ACACCGGGGT GCTGTTCAGC ACCCATGCCG GAAGCCAGAA AGGCCGCGAA 3890

CTGACAGAGA ACCCCTGGGC CTCGGGGACG CTGTATTGGC GCGAAACCAG CCAGCAGATC 3950

ATCCTCAATG GCCAGGCCGT GCGCATGCCG GATGCCAAGG CTGACGAGGC CTGGTTGAAG 4010

CGCCCTTATG CCACGCATCC GATGTCATCG GTGTCTCGCC AGAGTGAAGA ACTCAAGGAT 4070

GTTCAAGCCA TGCGCAACGC CGCCAGGGAA CTGGCCGAGG TTCAAGGTCC GCTGCCGCGT 4130

CCCGAGGGTT ATTGCGTGTT TGAGTTACGG CTTGAATCGC TGGAGTTCTG GGGTAACGGC 4190

GAGGAGCGCC TGCATGAACG CTTGCGCTAT GACCGCAGCG CTGAAGGCTG GAAACATCGC 4250

CGGTTACAGC CATAGGGTCC CGCGATAAAC ATGCTTTGAA GTGCCTGGCT GCTCCAGCTT 4310

CGAACTCATT GCGCAAACTT CAACACTTAT GACACCCGGT CAACATGAGA AAAGTCCAGA 4370

TGCGAAAGAA CGCGTATTCG AAATACCAAA CAGAGAGTCC GGATCACCAA AGTGTGTAAC 4430

GACATTAACT CCTATCTGAA TTTTATAGTT GCTCTAGAAC GTTGTCCTTG ACCCAGCGAT 4490

AGACATCGGG CCAGAACCTA CATAAACAAA GTCAGACATT ACTGAGGCTG CTACCATGCT 4550

AGATTTTCAA AACAAGCGTA AATATCTGAA AAGTGCAGAA TCCTTCAAAG CTT 4603

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 456 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Thr Gly Ile Pro Ser Ile Val Pro Tyr Ala Leu Pro Thr Asn Arg
1 5 10 15

Asp Leu Pro Val Asn Leu Ala Gln Trp Ser Ile Asp Pro Glu Arg Ala
20 25 30

Val Leu Leu Val His Asp Met Gln Arg Tyr Phe Leu Arg Pro Leu Pro

35 40 45

Asp Ala Leu Arg Asp Glu Val Val Ser Asn Ala Ala Arg Ile Arg Gln
50 55 60

Trp Ala Ala Asp Asn Gly Val Pro Val Ala Tyr Thr Ala Gln Pro Gly
65 70 75 80

Ser Met Ser Glu Glu Gln Arg Gly Leu Leu Lys Asp Phe Trp Gly Pro
85 90 95

Gly Met Lys Ala Ser Pro Ala Asp Arg Glu Val Val Gly Ala Leu Thr
100 105 110

Pro Lys Pro Gly Asp Trp Leu Leu Thr Lys Trp Arg Tyr Ser Ala Phe
115 120 125

Phe Asn Ser Asp Leu Leu Glu Arg Met Arg Ala Asn Gly Arg Asp Gln
130 135 140

Leu Ile Leu Cys Gly Val Tyr Ala His Val Gly Val Leu Ile Ser Thr
145 150 155 160

Val Asp Ala Tyr Ser Asn Asp Ile Gln Pro Phe Leu Val Ala Asp Ala
165 170 175

Ile Ala Asp Phe Ser Lys Glu His His Trp Met Pro Ser Asn Thr Pro
180 185 190

Pro Ala Val Ala Pro Cys His His His Arg Arg Gly Gly Ala Met Ser
195 200 205

Gln Thr Ala Ala His Leu Met Glu Arg Ile Leu Gln Pro Ala Pro Glu
210 215 220

Pro Phe Ala Leu Leu Tyr Arg Pro Glu Ser Ser Gly Pro Gly Leu Leu
225 230 235 240

Asp Val Leu Ile Gly Glu Met Ser Glu Pro Gln Val Leu Ala Asp Ile
245 250 255

Asp Leu Pro Ala Thr Ser Ile Gly Ala Pro Arg Leu Asp Val Leu Ala
260 265 270

Leu Ile Pro Tyr Arg Gln Ile Ala Glu Arg Gly Phe Glu Ala Val Asp
275 280 285

Asp Glu Ser Pro Leu Leu Ala Met Asn Ile Thr Glu Gln Gln Ser Ile
290 295 300

Ser Ile Glu Arg Leu Leu Gly Met Leu Pro Asn Val Pro Ile Gln Leu
305 310 315 320

Asn Ser Glu Arg Phe Asp Leu Ser Asp Ala Ser Tyr Ala Glu Ile Val
325 330 335

Ser Gln Val Ile Ala Asn Glu Ile Gly Ser Gly Glu Gly Ala Asn Phe
340 345 350

Val Ile Lys Arg Thr Phe Leu Ala Glu Ile Ser Glu Tyr Gly Pro Ala
355 360 365

Ser Ala Leu Ser Phe Phe Arg His Leu Leu Glu Arg Glu Lys Gly Ala
370 375 380

Tyr Trp Thr Phe Ile Ile His Thr Gly Ser Arg Thr Phe Val Gly Ala
385 390 395 400

Ser Pro Glu Arg His Ile Ser Ile Lys Asp Gly Leu Ser Val Met Asn
405 410 415

Pro Ile Ser Gly Thr Tyr Arg Tyr Pro Pro Ala Gly Pro Asn Leu Ser
420 425 430

Glu Val Met Asp Phe Leu Ala Asp Arg Lys Glu Ala Asp Glu Leu Tyr
435 440 445

Met Val Val Asp Glu Glu Leu *
450 455

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 388 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Met Met Ala Arg Ile Cys Glu Asp Gly Gly His Val Leu Gly Pro Tyr
1 5 10 15

Leu Lys Glu Met Ala His Leu Ala His Thr Glu Tyr Phe Ile Glu Gly
20 25 30

Lys Thr His Arg Asp Val Arg Glu Ile Leu Arg Glu Thr Leu Phe Ala
35 40 45

Pro Thr Val Thr Gly Ser Pro Leu Glu Ser Ala Cys Arg Val Ile Gln
50 55 60

Arg Tyr Xaa Pro Gln Gly Arg Ala Tyr Tyr Ser Gly Met Ala Ala Leu
65 70 75 80

Ile Gly Ser Asp Gly Lys Gly Gly Arg Ser Leu Asp Ser Ala Ile Leu
85 90 95

Ile Arg Thr Ala Asp Ile Asp Asn Ser Gly Glu Val Arg Ile Ser Val
100 105 110

Gly Ser Thr Ile Val Arg His Ser Asp Pro Met Thr Glu Ala Ala Glu
115 120 125

Ser Arg Ala Lys Ala Thr Gly Leu Ile Ser Ala Leu Lys Asn Gln Ala
130 135 140

Pro Ser Arg Phe Gly Asn His Leu Gln Val Arg Ala Ala Leu Ala Ser
145 150 155 160

Arg Asn Ala Tyr Val Ser Asp Phe Trp Leu Met Asp Ser Gln Gln Arg
165 170 175

Glu Gln Ile Gln Ala Asp Phe Ser Gly Arg Gln Val Leu Ile Val Asp
180 185 190

Ala Glu Asp Thr Phe Thr Ser Met Ile Ala Lys Gln Leu Arg Ala Leu
195 200 205

Gly Leu Val Val Thr Val Cys Ser Phe Ser Asp Glu Tyr Ser Phe Glu
210 215 220

Gly Tyr Asp Leu Val Ile Met Gly Pro Gly Pro Gly Asn Pro Ser Glu
225 230 235 240

Val Gln Gln Pro Lys Ile Asn His Leu His Val Ala Ile Arg Ser Leu
245 250 255

Leu Ser Gln Gln Arg Pro Phe Leu Ala Val Cys Leu Ser His Gln Val
260 265 270

Leu Ser Leu Cys Leu Gly Leu Glu Leu Gln Arg Lys Ala Ile Pro Asn
275 280 285

Gln Gly Val Gln Lys Gln Ile Asp Leu Phe Gly Asn Val Glu Arg Val
290 295 300

Gly Phe Tyr Asn Thr Phe Ala Ala Gln Ser Ser Ser Asp Arg Leu Asp
305 310 315 320

Ile Asp Gly Ile Gly Thr Val Glu Ile Ser Arg Asp Ser Glu Thr Gly
325 330 335

Glu Val His Ala Leu Arg Gly Pro Ser Phe Ala Ser Met Gln Phe His
340 345 350

Ala Glu Ser Leu Leu Thr Gln Glu Gly Pro Arg Ile Ile Ala Asp Leu
355 360 365

Leu Arg His Ala Leu Ile His Thr Pro Val Glu Asn Asn Ala Ser Ala
370 375 380

Ala Gly Arg *

385

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 279 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met His His Tyr Val Ile Ile Asp Ala Phe Ala Ser Val Pro Leu Glu
1 5 10 15

Gly Asn Pro Val Ala Val Phe Phe Asp Ala Asp Asp Leu Ser Ala Glu
20 25 30

Gln Met Gln Arg Ile Ala Arg Glu Met Asn Leu Ser Glu Thr Thr Phe
35 40 45

Val Leu Lys Pro Arg Asn Cys Gly Asp Ala Leu Ile Arg Ile Phe Thr
50 55 60

Pro Val Asn Glu Leu Pro Phe Ala Gly His Pro Leu Leu Gly Thr Asp
65 70 75 80

Ile Ala Leu Gly Ala Arg Thr Asp Asn His Arg Leu Phe Leu Glu Thr
85 90 95

Gln Met Gly Thr Ile Ala Phe Glu Leu Glu Arg Gln Asn Gly Ser Val
100 105 110

Ile Ala Ala Ser Met Asp Gln Pro Ile Pro Thr Trp Thr Ala Leu Gly
115 120 125

Arg Asp Ala Glu Leu Leu Lys Ala Leu Gly Ile Ser Asp Ser Thr Phe
130 135 140

Pro Ile Glu Ile Tyr His Asn Gly Pro Arg His Val Phe Val Gly Leu
145 150 155 160

Pro Ser Ile Ala Ala Leu Ser Ala Leu His Pro Asp His Arg Ala Leu
165 170 175

Tyr Ser Phe His Asp Met Ala Ile Asn Cys Phe Ala Gly Ala Gly Arg
180 185 190

Arg Trp Arg Ser Arg Met Phe Ser Pro Ala Tyr Gly Val Val Glu Asp
195 200 205

Ala Xaa Thr Gly Ser Ala Ala Gly Pro Leu Ala Ile His Leu Ala Arg
210 215 220

```
His Gly Gln Ile Glu Phe Gly Gln Gln Ile Glu Ile Leu Gln Gly Val
225                 230                 235                 240

Glu Ile Gly Arg Pro Ser Leu Met Phe Ala Arg Ala Glu Gly Arg Ala
                245                 250                 255

Asp Gln Leu Thr Arg Val Glu Val Ser Gly Asn Gly Ile Thr Phe Gly
            260                 265                 270

Arg Gly Thr Ile Val Leu  *
        275
```

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1007 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..669
(D) OTHER INFORMATION: /gene= "phz4"
/label= ORF4
/note= "This DNA sequence is repeated from SEQ ID NO:17 so that the overlapping ORF4 may be separately translated"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```
ATG AAC AGT TCA GTA CTA GGC AAG CCG CTG TTG GGT AAA GGC ATG TCG        48
Met Asn Ser Ser Val Leu Gly Lys Pro Leu Leu Gly Lys Gly Met Ser
  1               5                  10                  15

GAA TCG CTG ACC GGC ACA CTG GAT GCG CCG TTC CCC GAG TAC CAG AAG        96
Glu Ser Leu Thr Gly Thr Leu Asp Ala Pro Phe Pro Glu Tyr Gln Lys
             20                  25                  30

CCG CCT GCC GAT CCC ATG AGC GTG CTG CAC AAC TGG CTC GAA CGC GCA       144
Pro Pro Ala Asp Pro Met Ser Val Leu His Asn Trp Leu Glu Arg Ala
         35                  40                  45

CGC CGC GTG GGC ATC CGC GAA CCC CGT GCG CTG GCG CTG GCC ACG GCT       192
Arg Arg Val Gly Ile Arg Glu Pro Arg Ala Leu Ala Leu Ala Thr Ala
```

50 55 60

GAC AGC CAG GGC CGG CCT TCG ACA CGC ATC GTG GTG ATC AGT GAG ATC 240
Asp Ser Gln Gly Arg Pro Ser Thr Arg Ile Val Val Ile Ser Glu Ile
65 70 75 80

AGT GAC ACC GGG GTG CTG TTC AGC ACC CAT GCC GGA AGC CAG AAA GGC 288
Ser Asp Thr Gly Val Leu Phe Ser Thr His Ala Gly Ser Gln Lys Gly
85 90 95

CGC GAA CTG ACA GAG AAC CCC TGG GCC TCG GGG ACG CTG TAT TGG CGC 336
Arg Glu Leu Thr Glu Asn Pro Trp Ala Ser Gly Thr Leu Tyr Trp Arg
100 105 110

GAA ACC AGC CAG CAG ATC ATC CTC AAT GGC CAG GCC GTG CGC ATG CCG 384
Glu Thr Ser Gln Gln Ile Ile Leu Asn Gly Gln Ala Val Arg Met Pro
115 120 125

GAT GCC AAG GCT GAC GAG GCC TGG TTG AAG CGC CCT TAT GCC ACG CAT 432
Asp Ala Lys Ala Asp Glu Ala Trp Leu Lys Arg Pro Tyr Ala Thr His
130 135 140

CCG ATG TCA TCG GTG TCT CGC CAG AGT GAA GAA CTC AAG GAT GTT CAA 480
Pro Met Ser Ser Val Ser Arg Gln Ser Glu Glu Leu Lys Asp Val Gln
145 150 155 160

GCC ATG CGC AAC GCC GCC AGG GAA CTG GCC GAG GTT CAA GGT CCG CTG 528
Ala Met Arg Asn Ala Ala Arg Glu Leu Ala Glu Val Gln Gly Pro Leu
165 170 175

CCG CGT CCC GAG GGT TAT TGC GTG TTT GAG TTA CGG CTT GAA TCG CTG 576
Pro Arg Pro Glu Gly Tyr Cys Val Phe Glu Leu Arg Leu Glu Ser Leu
180 185 190

GAG TTC TGG GGT AAC GGC GAG GAG CGC CTG CAT GAA CGC TTG CGC TAT 624
Glu Phe Trp Gly Asn Gly Glu Glu Arg Leu His Glu Arg Leu Arg Tyr
195 200 205

GAC CGC AGC GCT GAA GGC TGG AAA CAT CGC CGG TTA CAG CCA TAGGGTCCCG 676
Asp Arg Ser Ala Glu Gly Trp Lys His Arg Arg Leu Gln Pro
210 215 220

CGATAAACAT GCTTTGAAGT GCCTGGCTGC TCCAGCTTCG AACTCATTGC GCAAACTTCA 736

ACACTTATGA CACCCGGTCA ACATGAGAAA AGTCCAGATG CGAAAGAACG CGTATTCGAA 796

ATACCAAACA GAGAGTCCGG ATCACCAAAG TGTGTAACGA CATTAACTCC TATCTGAATT 856

TTATAGTTGC TCTAGAACGT TGTCCTTGAC CCAGCGATAG ACATCGGGCC AGAACCTACA 916

TAAACAAAGT CAGACATTAC TGAGGCTGCT ACCATGCTAG ATTTTCAAAA CAAGCGTAAA 976

TATCTGAAAA GTGCAGAATC CTTCAAAGCT T 1007

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```
Met Asn Ser Ser Val Leu Gly Lys Pro Leu Leu Gly Lys Gly Met Ser
  1               5                  10                  15

Glu Ser Leu Thr Gly Thr Leu Asp Ala Pro Phe Pro Glu Tyr Gln Lys
             20                  25                  30

Pro Pro Ala Asp Pro Met Ser Val Leu His Asn Trp Leu Glu Arg Ala
         35                  40                  45

Arg Arg Val Gly Ile Arg Glu Pro Arg Ala Leu Ala Leu Ala Thr Ala
     50                  55                  60

Asp Ser Gln Gly Arg Pro Ser Thr Arg Ile Val Val Ile Ser Glu Ile
 65                  70                  75                  80

Ser Asp Thr Gly Val Leu Phe Ser Thr His Ala Gly Ser Gln Lys Gly
                 85                  90                  95

Arg Glu Leu Thr Glu Asn Pro Trp Ala Ser Gly Thr Leu Tyr Trp Arg
            100                 105                 110

Glu Thr Ser Gln Gln Ile Ile Leu Asn Gly Gln Ala Val Arg Met Pro
        115                 120                 125

Asp Ala Lys Ala Asp Glu Ala Trp Leu Lys Arg Pro Tyr Ala Thr His
    130                 135                 140

Pro Met Ser Ser Val Ser Arg Gln Ser Glu Glu Leu Lys Asp Val Gln
145                 150                 155                 160

Ala Met Arg Asn Ala Ala Arg Glu Leu Ala Glu Val Gln Gly Pro Leu
                165                 170                 175

Pro Arg Pro Glu Gly Tyr Cys Val Phe Glu Leu Arg Leu Glu Ser Leu
            180                 185                 190

Glu Phe Trp Gly Asn Gly Glu Glu Arg Leu His Glu Arg Leu Arg Tyr
        195                 200                 205

Asp Arg Ser Ala Glu Gly Trp Lys His Arg Arg Leu Gln Pro
    210                 215                 220
```

What is claimed is:

1. An isolated DNA molecule encoding one or more polypeptides required for the biosynthesis of an antipathogenic substance (APS) in a heterologous host, wherein said APS is selected from the group consisting of pyrrolnitrin and soraphen.

2. The isolated DNA molecule of claim 1, wherein said APS is pyrrolnitrin and said polypeptide is selected from the group consisting of SEQ ID Nos. 2-5.

3. The isolated DNA molecule of claim 1, wherein said APS is pyrrolnitrin and said DNA molecule has the sequence set forth in SEQ ID No. 1.

4. The isolated DNA molecule of claim 1, wherein said APS is soraphen and said DNA molecule has the sequence set forth in SEQ ID No. 6.

5. The DNA molecule according to any one of claims 1 to 4 engineered to form part of a plant genome.

6. An expression vector comprising the isolated DNA molecule of claim 1 wherein said vector is capable of expressing one or more polypeptides encoded by said DNA molecule in a host cell.

7. A heterologous host transformed with an expression vector comprising the isolated DNA molecule of claim 1, wherein said host is selected from the group consisting of a bacterium, a fungus, a yeast and a plant.

8. The heterologous host of claim 7, wherein said host is a plant.

9. A host capable of synthesizing an antipathogenic substance not naturally occurring in said host.

10. The host of claim 9, wherein said antipathogenic substance is selected from the group consisting of a carbohydrate containing antibiotic, a peptide antibiotic, a heterocyclic

antibiotic containing nitrogen, a heterocyclic antibiotic containing oxygen, a heterocyclic antibiotic containing nitrogen and oxygen, a polyketide, a macrocyclic lactone, and a quinone.

11. The host of claim 10, wherein said peptide antibiotic is rhizocticin.

12. The host of claim 10, wherein said carbohydrate containing antibiotic is an aminoglycoside.

13. The host of claim 10, wherein said antipathogenic substance is a heterocyclic antibiotic containing nitrogen.

14. The host of claim 13, wherein said heterocyclic antibiotic containing nitrogen is selected from the group consisting of phenazine and pyrrolnitrin.

15. The host of claim 10, wherein said antipathogenic substance is a polyketide.

16. The host of claim 15, wherein said polyketide is soraphen.

17. The host of claim 9, wherein said antipathogenic substance is resorcinol.

18. The host of claim 9, wherein said antipathogenic substance is a methoxyacrylate.

19. The host of claim 18, wherein said methoxyacrylate is strobilurin B.

20. The host of claim 9, wherein said host is selected from the group consisting of a plant, a bacterium, a yeast and a fungus.

21. The host of claim 20, wherein said host is a plant.

22. The host of claim 21, wherein said host is a hybrid plant.

23. Propagating material of a host according to claim 21 or 22 treated with a protectant coating.

24. Propagating material according to claim 23, comprising a preparation selected from the group consisting of herbicides, insecticides, fungicides, bactericides, nematicides, molluscicides or mixtures thereof.

25. Propagating material according to claim 23 or 24 characterized in that it consists of seed.

26. The host of claim 20, wherein said host is a biocontrol agent.

27. The host of claim 20, wherein said host is a plant colonizing organism.

28. The host of claim 20, wherein said host is suitable for producing large quantities of said APS.

29. A host capable of synthesizing enhanced amounts of an antipathogenic substance naturally occurring in said host, wherein said host is transformed with one or more DNA molecules collectively encoding the complete set of polypeptides required to synthesize said antipathogenic substance.

30. A method for protecting a plant against a phytopathogen comprising transforming said plant with one or more vectors collectively capable of expressing all of the polypeptides necessary to produce an anti-phytopathogenic substance in said plant in amounts which inhibit said phytopathogen.

31. A method for protecting a plant against a phytopathogen comprising treating said plant with a biocontrol agent transformed with one or more vectors collectively capable of expressing all of the polypeptides necessary to produce an anti-phytopathogenic substance in amounts which inhibit said phytopathogen.

32. A method for protecting a plant against a phytopathogen comprising applying to said plant a composition comprising an anti-phytopathogenic substance in amounts which inhibit

said phytopathogen, wherein said anti-phytopathogenic substance is obtained from the host of claim 28.

33. A method for producing large quantities of an antipathogenic substance (APS) of uniform chirality comprising

(a) transforming a host with one or more vectors collectively capable of expressing all of the polypeptides necessary to produce said APS in said host;

(b) growing said host under conditions which allow production of said APS; and

(c) collecting said APS from said host.

34. A composition comprising an antipathogenic substance (APS) of uniform chirality produced by the method of claim 33.

35. A method for identifying and isolating a gene from a microorganism required for the biosynthesis of an antipathogenic substance (APS), wherein the expression of said gene is under the control of a regulator of the biosynthesis of said APS, said method comprising

(a) cloning a library of genetic fragments from said microorganism into a vector adjacent to a promoterless reporter gene in a vector such that expression of said reporter gene can occur only if promoter function is provided by the cloned fragment;

(b) transforming the vectors generated from step (a) into a suitable host;

(c) identifying those transformants from step (b) which express said reporter gene only in the presence of said regulator; and

(d) identifying and isolating the DNA fragment operably linked to the genetic fragment from said microorganism present in the transformants identified in step (c);

wherein said DNA fragment isolated and identified in step (d) encodes one or more polypeptides required for the biosynthesis of said APS.

36. An isolated polypeptide required for the biosynthesis of an antipathogenic substance (APS) in a heterologous host, wherein said APS is selected from the group consisting of pyrrolnitrin and soraphen.

37. The isolated polypeptide of claim 36, wherein said APS is pyrrolnitrin and said polypeptide is selected from the group consisting of SEQ ID Nos. 2-5.

38. The isolated polypeptide claim 36, wherein said APS is pyrrolnitrin and said polypeptide is encoded by the nucleotide sequence set forth in SEQ ID No. 1.

39. The isolated polypeptide of claim 36, wherein said APS is soraphen and said polypeptide is encoded by the nucleotide sequence set forth in SEQ ID No. 6.

40. Use of a DNA molecule according to claim 1 for genetically engineering a host organism to express said antipathogenic substance.

41. Use according to claim 40, wherein said host is selected from the group consisting of a plant, a bacterium, a yeast and a fungus.

42. Use according to claim 40, wherein the antipathogenic substance expressed does not naturally occur in said host.

43. Use according to claim 40, wherein increased amounts of the antipathogenic substance naturally occurring in said host are produced.

44. Use of the host according to claim 7 for protecting a plant against a phytopathogen.

45. Use of the composition according to claim 34 for protecting a plant against a phytopathogen.

46. Use of the DNA molecule according to claim 5 to transfer the ability to express an antipathogenic molecule from a parent plant to its progeny.

pCIB169 Restriction Map

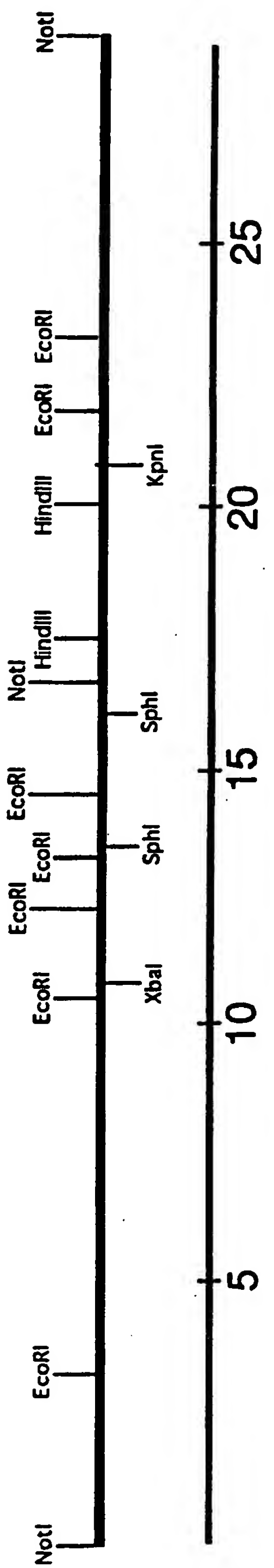


Fig. 1

Functional Map of the Pyrrolnitrin Gene Region of MOCG134

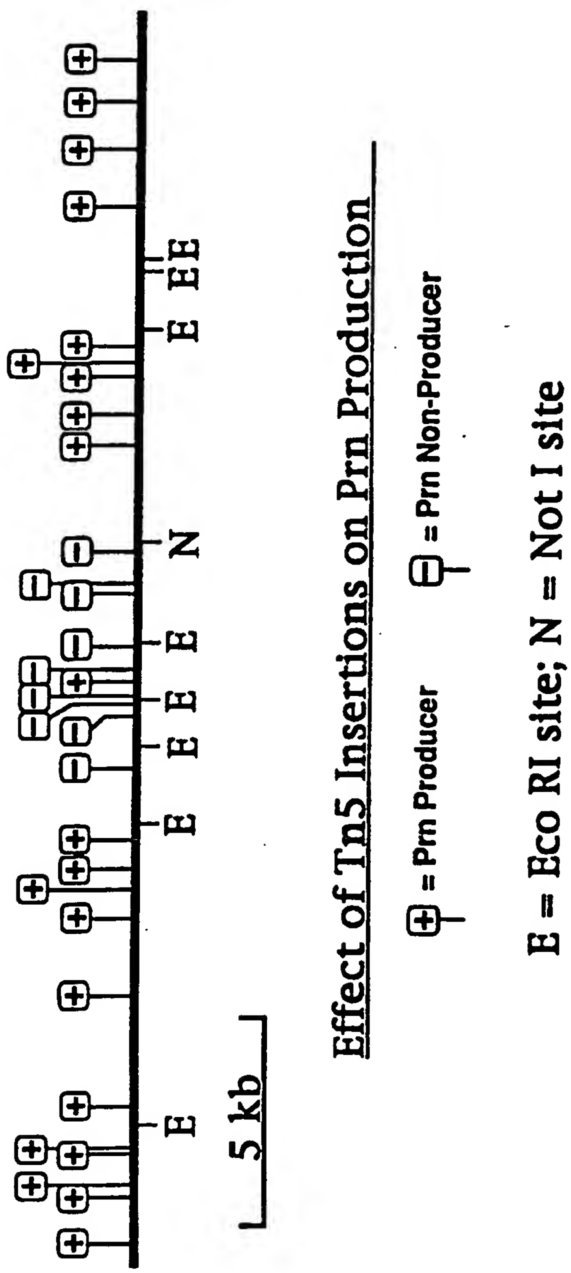


Fig. 2

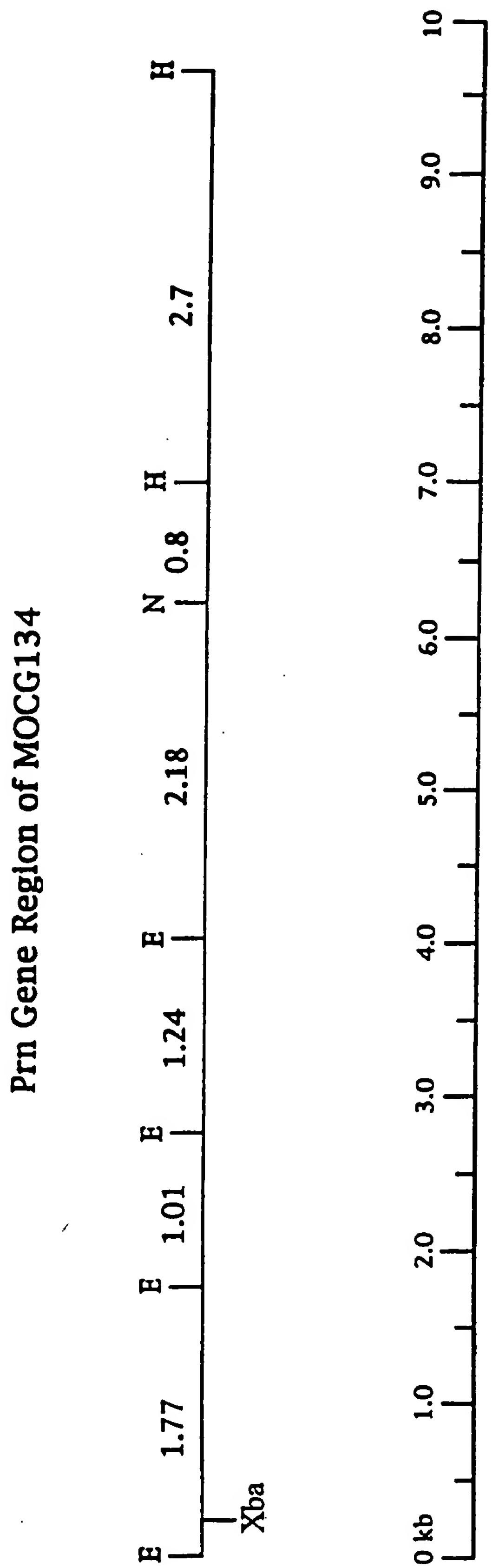

Prn Gene Region of MOCG134
E
Xba
1.77
E
1.01
E
1.24
E
2.18
N
0.8
H
2.7
H
0 kb
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10

Fig. 3

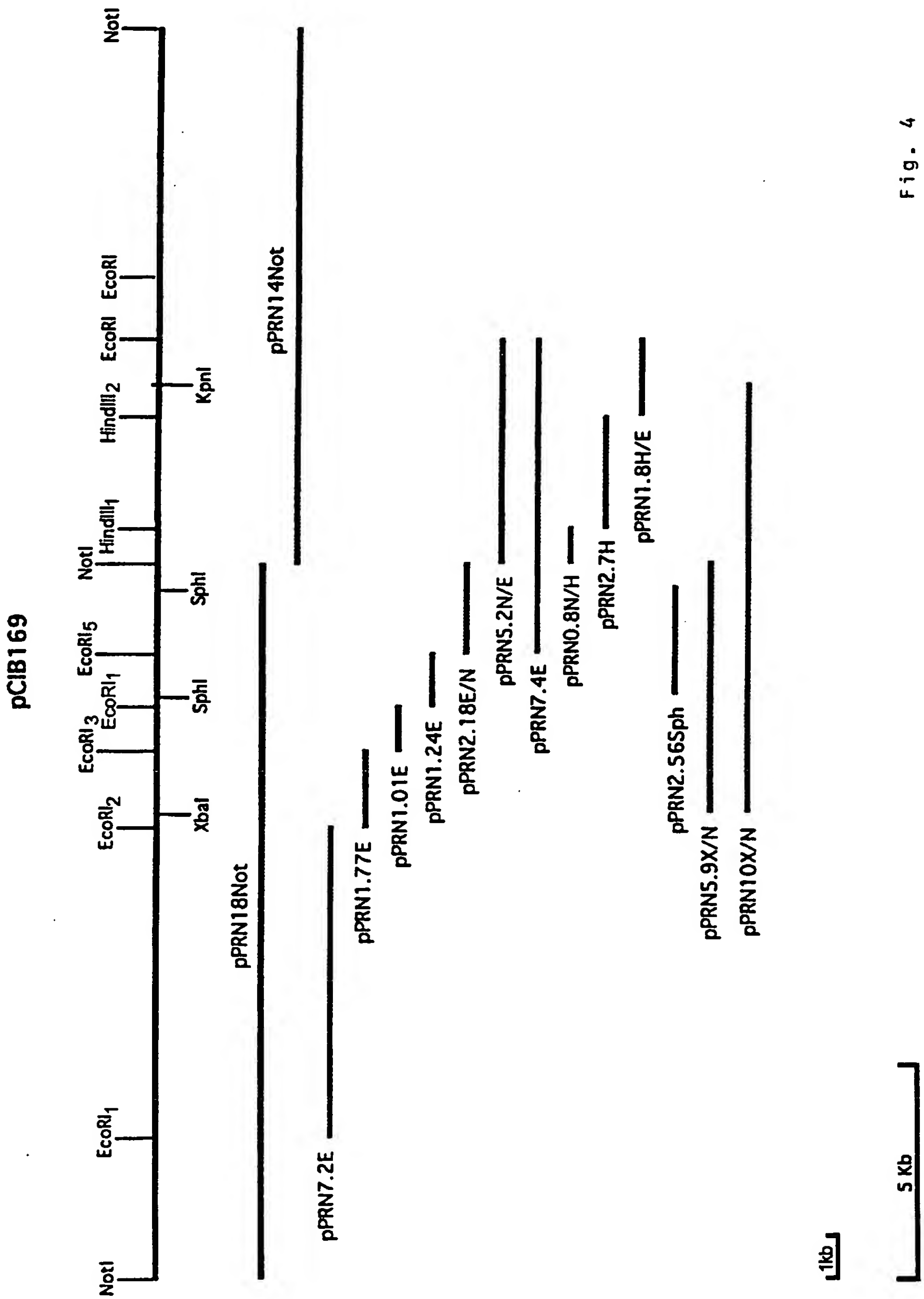

pCIB169
NotI
EcoRI1
EcoRI2
XbaI
EcoRI3
EcoRI1
SphI
EcoRI5
SphI
NotI
HindIII1
HindIII2
KpnI
EcoRI
EcoRI
NotI
pPRN18Not
pPRN14Not
pPRN7.2E
pPRN1.77E
pPRN1.01E
pPRN1.24E
pPRN2.18E/N
pPRN5.2N/E
pPRN7.4E
pPRN0.8N/H
pPRN2.7H
pPRN1.8H/E
pPRN2.56Sph
pPRN5.9X/N
pPRN10X/N
1kb
5 Kb

Fig. 4

Prn Gene Region of MOCG134

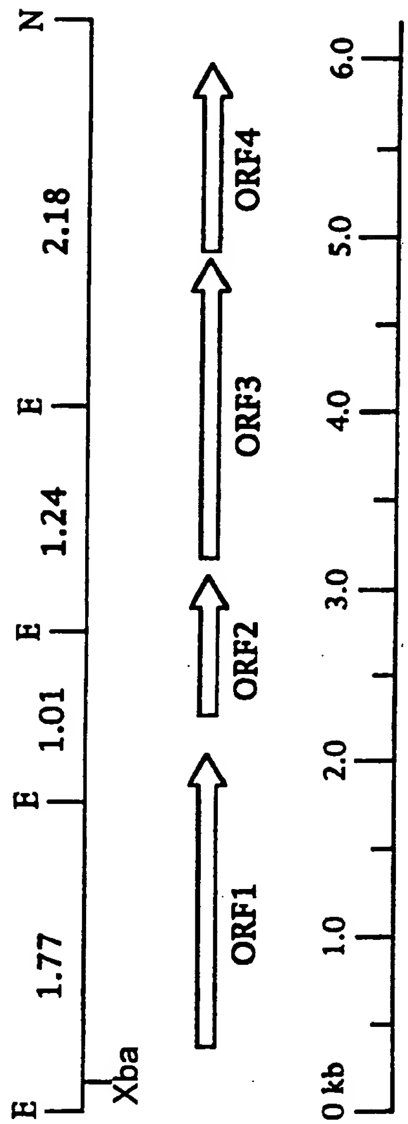


Fig. 5

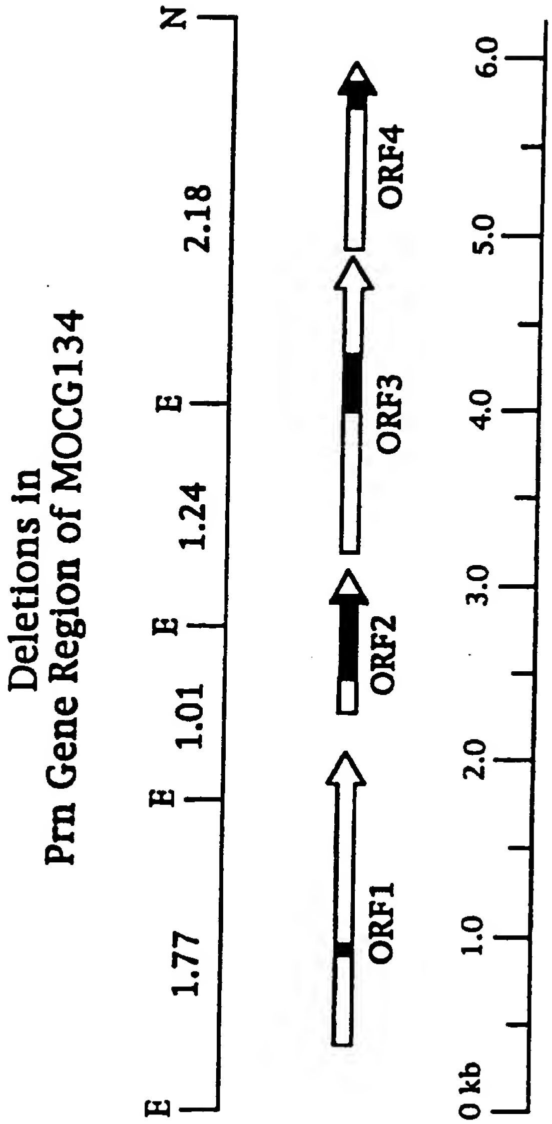

Deletions in
Prn Gene Region of MOCG134
E
E
E
E
N
1.77
1.01
1.24
2.18
ORF1
ORF2
ORF3
ORF4
0 kb
1.0
2.0
3.0
4.0
5.0
6.0

Fig. 6

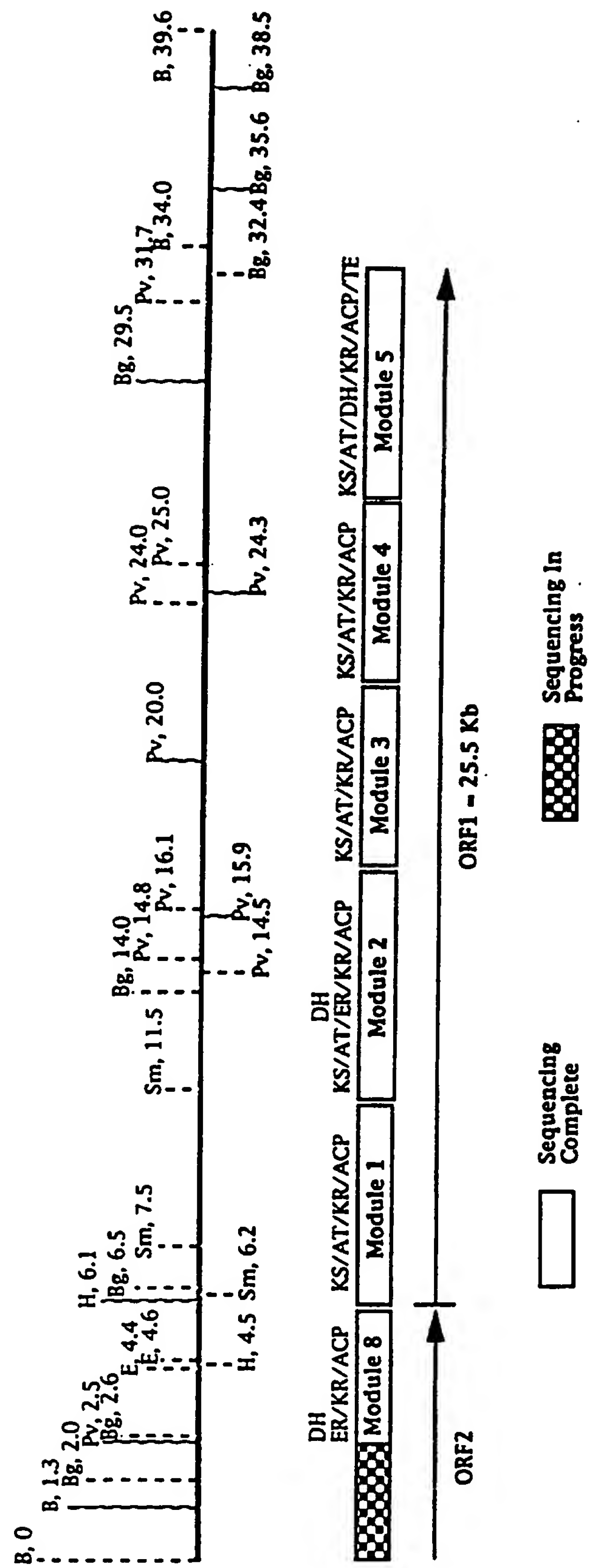

B, 0
B, 1.3
Bg, 2.0
Pv, 2.5
Bg, 2.6
E, 4.4
E, 4.6
H, 6.1
Bg, 6.5
Sm, 7.5
H, 4.5
Sm, 6.2
Sm, 11.5
Bg, 14.0
Pv, 14.8
Pv, 16.1
Pv, 14.5
Pv, 15.9
Pv, 20.0
Pv, 24.0
Pv, 25.0
Pv, 24.3
Bg, 29.5
Pv, 31.7
B, 34.0
Bg, 32.4
Bg, 35.6
B, 39.6
Bg, 38.5
DH
ER/KR/ACP
Module 8
KS/AT/KR/ACP
Module 1
DH
KS/AT/ER/KR/ACP
Module 2
KS/AT/KR/ACP
Module 3
KS/AT/KR/ACP
Module 4
KS/AT/DH/KR/ACP/TE
Module 5
ORF2
ORF1 = 25.5 Kb
Sequencing Complete
Sequencing In Progress

Fig. 7

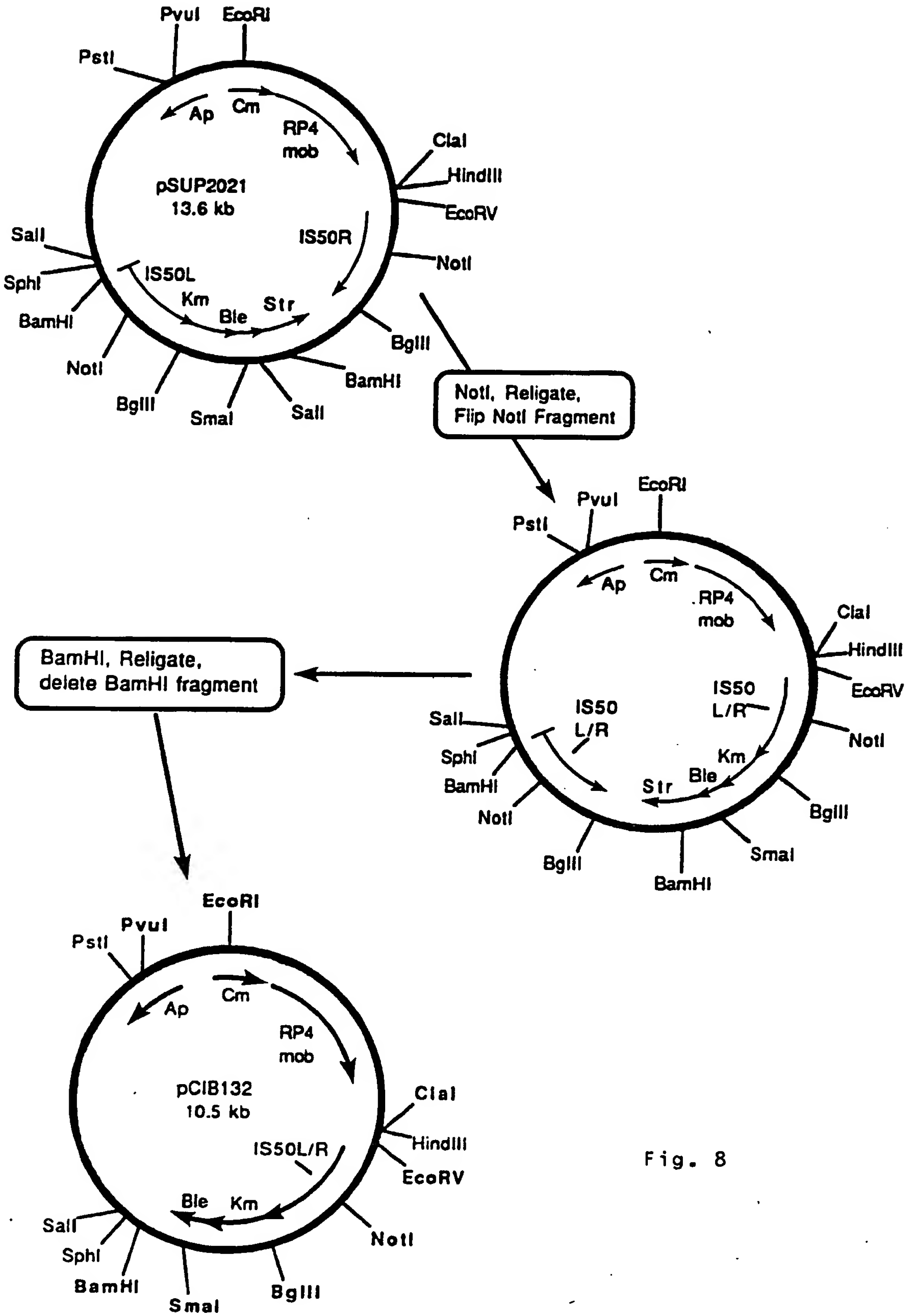

PvuI
EcoRI
PstI
Ap
Cm
RP4
mob
ClaI
HindIII
EcoRV
pSUP2021
13.6 kb
IS50R
SalI
SphI
BamHI
IS50L
Km
Ble
Str
NotI
BglII
NotI
BglII
SmaI
SalI
BamHI
NotI, Religate,
Flip NotI Fragment
EcoRI
PvuI
PstI
Ap
Cm
RP4
mob
ClaI
HindIII
EcoRV
BamHI, Religate,
delete BamHI fragment
IS50
L/R
IS50
L/R
SalI
SphI
BamHI
NotI
Km
Ble
Str
NotI
BglII
SmaI
BglII
BamHI
EcoRI
PvuI
PstI
Ap
Cm
RP4
mob
pCIB132
10.5 kb
ClaI
HindIII
EcoRV
IS50L/R
Ble
Km
SalI
SphI
BamHI
SmaI
BglII
NotI

Fig. 8

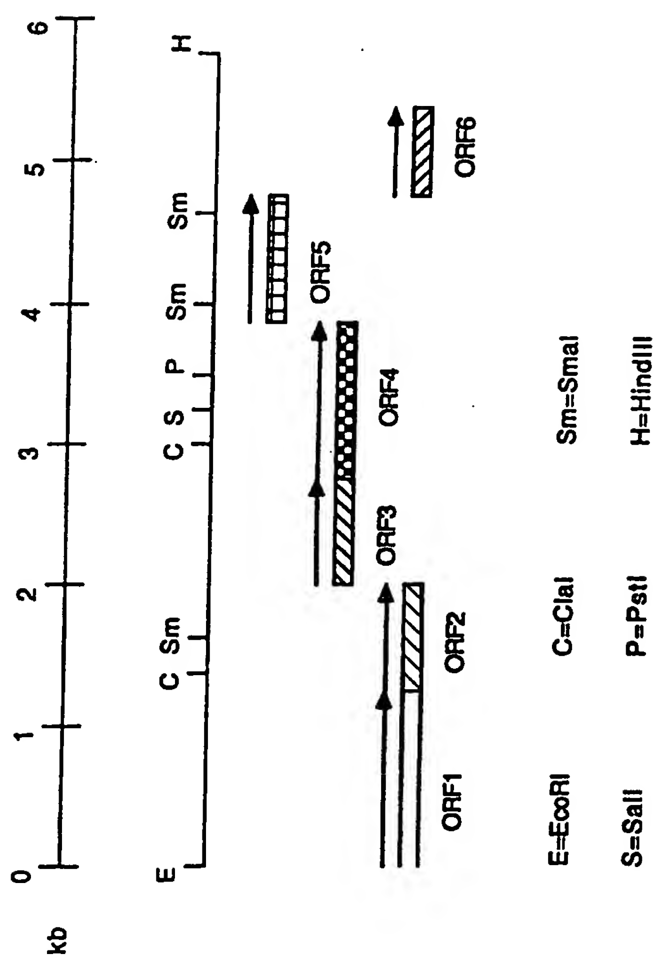

kb
0
1
2
3
4
5
6
E
C
Sm
C
S
P
Sm
Sm
H
ORF1
ORF2
ORF3
ORF4
ORF5
ORF6
E=EcoRI
S=SalI
C=ClaI
P=PstI
Sm=SmaI
H=HindIII

Fig. 9